Centre de recherche sur les transports (C.R.T.)
Centre for Research on Transportation

CRT-2006-13

Fleet Management for Advanced Intermodal Services

Final Report

by

Teodor Gabriel Crainic
Michel Gendreau
Ioana Bilegan

July 2006

Université de Montréal
C.P. 6128, succursale Centre-ville, Montréal QC H3C 3J7 Canada
Téléphone: (514) 343-7575 / Télécopieur: (514) 343-7121

crt@crt.umontreal.ca
www.crt.umontreal.ca

HEC

ÉCOLE
POLYTECHNIQUE
MONTRÉAL

Université
de Montréal

106-06871/

PUBLICATION
CRT-2006-13

Fleet Management for Advanced Intermodal Services

*Final Report**

by

Teodor Gabriel Crainic[1]
Michel Gendreau[2]
Ioana Bilegan[3]

* *Project RDA-01* - R&D Contribution Agreements Program INNOVATION through Partnership Intelligent Transportation Systems Research & Development Plan for Canada, Transport Canada.

1. Département de management et technologie, École des sciences de la gestion, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville, Montréal, Canada H3C 3P8 and Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7 (theo@crt.umontreal.ca)
2. Département d'informatique et de recherche opérationnelle and Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Canada H3C 3J7 (michelg@crt.umontreal.ca)
3. Université de Valenciennes LAMIH-ROI (istv2), Le Mont Houy F 59313 Valenciennes, France (ioana.bilegan@univ-valenciennes.fr)

Département d'informatique et de recherche opérationnelle - Publication #1280

Dépôt légal – Bibliothèque nationale du Québec
Bibliothèque nationale du Canada, 2006

Acknowledgments

This is the final report of the project RDA-01 of the ITS Bureau of Transport Canada. It corresponds to Milestone Task #5. The report has been prepared by Dr. Ioana Bilegan, Professor Teodor Gabriel Crainic, and Professor Michel Gendreau. The authors wish to thank Mr. Gordon Graham, CN, Mr. Raynald Ledoux, Transport Canada, and Mr. Luc Lefebvre, ministère des Transports du Québec, for their comments on an earlier version of the report.

This project has been 40% funded by Transport Canada through a contribution agreement, under Canada's Intelligent Transportation Systems R&D Plan "Innovation through Partnership". The Plan is part of the ITS component of the Strategic Highway Infrastructure Program.

Funding for this project has also been provided by the CN Chair in Intermodal Transportation (CRT, Université de Montréal), the Centre for Research on Transportation (CRT), and the team of university professors. Additional funding has been provided by the NSERC Industrial Research Chair in Logistics Management (École des sciences de la gestion, UQAM).

Remerciements

Ce rapport a été réalisé par Dr. Ioana Bilegan, Professeur Teodor Gabriel Crainic et Professeur Michel Gendreau. Les auteurs remercient M. Gordon Graham du CN, M. Raynald Ledoux de Transports Canada et M. Luc Lefebvre du ministère des Transports du Québec.

Ce projet a été financé à 40% par Transports Canada par le biais d'un accord de contribution dans le cadre du plan de R&D des systèmes de transport intelligents pour le Canada « Innover par l'établissement de partenariats ». Le projet fait partie du volet STI du Programme stratégique d'infrastructures routières. Le projet a aussi été financé par des contributions de la Chaire CN en transport intermodal (C.R.T., Université de Montréal), du Centre de recherche sur les transports (C.R.T.) et de l'équipe de professeurs. Des fonds supplémentaires ont été accordés par la Chaire de recherche industrielle du CRSNG en management logistique (École des sciences de la gestion, UQAM).

Abstract

This is the final report of the project undertook by a C.R.T. research team, in collaboration with CN and the support of Transport Canada, on the topic of intelligent intermodal freight transportation systems. The project was part of the effort to improve intermodal transportation in Canada and addressed an innovative rail intermodal service mode based on full-asset utilization principles, strict schedule, and advance bookings. The report presents a comprehensive view of the work performed, the results achieved, as well as the perspectives opened, the challenges ahead, and the contemplated plans to address them.

Keywords: Intelligent Transportation Systems, intermodal transportation, rail transportation, planning and operation models, forecast, simulation.

Résumé

Ce document représente le rapport final du projet réalisé par une équipe du C.R.T., en collaboration avec le CN et le support de Transports Canada, sur la problématique des systèmes intelligents de transport intermodaux de marchandises. Le projet visait à faire une contribution à l'amélioration du transport intermodal au Canada et avait pour principal objet d'intérêt un type innovateur de service de transport intermodal par rail basé sur des principes d'utilisation maximale des ressources, un horaire strict, ainsi qu'un système de réservations. Le rapport présente une vue d'ensemble du travail accompli et des résultats obtenus, ainsi qu'une discussion des perspectives et défis à venir.

Mots-clés : Systèmes de transport intelligents, systèmes de transport intermodal, transport ferroviaire, modèles de planification et d'exploitation, prévisions, simulation.

Introduction

The final report of the project *Fleet Management for Advanced Intermodal Services* presents a comprehensive view of the work performed, the results achieved, as well as the perspectives opened, the challenges ahead, and the contemplated plans to address them.

The report is structured as follows. Section I recalls the context, scope, and goals of the project, while Section II presents the project team and management structure. The project timeline and general methodology is recalled in Section III. Section IV presents a review of relevant literature. Section V concerns data issues: collection, understanding, and processing of data for the project requirements. Sections VI, VII, and VIII present the work performed and the results achieved for the three main components of the project: forecast of container-release times, development of a first-generation simulation tool for assessment and evaluation, and intermodal service design, respectively. Section IX summarizes the findings and results of the project, discusses challenges ahead, and proposes a roadmap to develop the methods and instruments required to address them. A list of references follows.

Annex A presents a mathematical formulation of the service-adjustment model described in Section VIII. Annex B gives details about the technical specification for the development of the train service simulator.

I. Project Scope and Goals

In a general sense, freight ***intermodal transportation*** refers to the movement of cargo using more than one transportation mode for its journey and requiring processing (transshipment, classification, consolidation, etc.) at associated intermodal terminals. A significant proportion of such intermodal, or multi-modal, transportation is performed by using containers. Intermodal transportation is therefore often equated to container transportation and carriers operating container-based transportation services are identified as intermodal carriers. This is particularly the case in North America where railways have created intermodal divisions focused on the efficient container transportation for the local and international markets.

Intermodal transportation plays a vital role in today's economy, particularly for a country like Canada that depends heavily on its importing and exporting activities, has the United States as a (the) major commercial partner and thus serves as an entry and exit gate of containers for North America, and is geographically very large, which implies long transportation distances for most cargo flow. The performance of transportation services, of intermodal transportation in particular, directly affects the operational (e.g., the just-in-time production and delivery paradigm) and economic efficiency of most industrial sectors and logistics networks (traditionally also know as logistics chains or value chains) and, thus, the availability, quality, and price of goods and services offered to the Canadian population or exported.

Intermodal transportation is, by definition, at least the use of two transportation modes with an intermodal transshipment operation. A classical intermodal chain has a first movement by truck, a transfer to rail or ship, a long-haul movement, followed by a transfer to truck for the final leg of the journey. In many cases, the long-haul movement involves more than one mode. Thus, for many Canadian imports and exports, goods leave or arrive in Canada by ship, the land part of the journey being performed by rail and truck.

Seaport container terminals play a central role in such a multi-modal chain. First, there is a complex series of operations containers must go through – ship loading/unloading, storage, intra-port movements, loading/unloading on rail, etc. – that generate delays and costs for all parties involved. Second, ports are gateways for Canada and the North-American continent. Consequently, custom and immigration controls are an integral part of the process import containers must go through. In the contemporary context of high security-awareness, custom and immigration controls add an increasingly significant amount to the complexity and resulting delay of port operations.

The performance of intermodal transportation thus directly depends on the performance of the key individual elements of the chain – e.g., rail carriers, ports, navigation companies, etc., – as well as on the quality of interactions between stakeholders regarding operations, information, and decisions. *Intelligent Transportation Systems* (*ITS*) models, methods, and technologies offer the means to address these challenges both for the individual carrier or port and for the integrated intermodal chain. To study and improve the performance of intermodal transportation one must therefore simultaneously focus on carrier service design and management and on the means to ensure smooth interactions and seamless exchanges between partners.

This project was part of the effort to improve intermodal transportation in Canada. The project addressed an innovative service mode proposed by ***CN***, one of the most important Canadian rail carriers serving the entire country for the national and international container traffic, as well as a significant proportion of our exchanges with the United States. The ***scheduled-with-bookings*** system operates regular scheduled services and requires customers to book space in advance. This new approach to operating intermodal rail services brings advantages for the carrier in terms of operating costs and asset utilization, as well as to the customers (once they get used to the new operating mode) in terms of increased reliability, regular and predictable service and, eventually, better price.

The service has been initially introduced in the Eastern network (the corridor Halifax, Montreal, Toronto, Chicago) and applied to the domestic traffic with great success. On each service (e.g., Halifax – Montreal – Toronto – Chicago), it operates fixed-length trains and blocks (e.g., Halifax-to-Montreal) according to a known schedule. The service has now been extended to the entire network for the domestic and export traffic. The next challenge is to extend this type of service to include the import international traffic, which represents a significant part of the total traffic.

The challenge rises from the uncertainties associated to the release of containers (and railcars) from ports. The variability associated with container-release times, due to the combined effect of ship travel time and arrival, as well as container terminal, port, and Customs operations, clashes with the regularity and rigidity of the intermodal rail service. This has blocked the extension of the booking system to the import international traffic.

The scope of the project therefore was to explore possible strategies for a better integration of the international traffic into the scheduled-with-bookings system. To achieve this goal, work was planned both on the *supply* and the *demand* components of the system. The demand side aimed toward a better understanding and modelling of the phenomena that determine the container-release times at port terminals, including the information exchanges among stakeholders. On the supply side, the issue was to propose and analyze possible booking systems for the international traffic or, alternatively, methods to adjust the transportation plan to demand while continuing to operate according to the full-asset-utilization policy of the rail carrier. The integration of these two components and the evaluation of various proposals required the development of simulation tools.

The objectives of the proposed work therefore were to:

1. Build a first-generation forecasting model for container-release times;
2. Propose and analyze strategies to better integrate international intermodal traffic and scheduled transportation services with bookings for freight;
3. Develop a first-generation simulation software to integrate and assess the proposed strategies;
4. Propose a blueprint for future developments.

II. Project Team and Management

The project has been undertook at the Centre for Research on Transportation (CRT), a multi-disciplinary and multi-institutional research center jointly sponsored by the Université de Montréal, HEC-Montréal, and École Polytechnique de Montréal.

The project team was composed of five university professors, CN personnel, a post-doctoral fellow, an analyst-programmer, and a few students. The principal university researchers were:

- Dr. Teodor Gabriel Crainic, **project manager**, Professor of Operations Research and NSERC Industrial Research Chair in Logistics Management, École des Sciences de la Gestion, Université du Québec à Montréal, and Director, Laboratory for Intelligent Transportation Systems, Centre for Research on Transportation.
- Professor Michel Gendreau, Département d'informatique et recherche opérationnelle (Computer Science and Operations Research) of Université de Montréal, and Director of the Centre for Research in Transportation.

Have also collaborated to the project:

- Dr. Robert Gagné, Professor of Economics at HEC Montréal.
- Dr. Claude Comtois, Professor of Geography, Université de Montréal.
- Dr. Brian Slack, Distinguished Professor Emeritus of Geography, Concordia University.

Several persons from CN have collaborated to the project and provided invaluable contributions in the form of knowledge, data, and guidance:

- Andrew Fuller, Director, Intermodal Product
- Gordon Graham, Network Manager Intermodal Commercial Operations

- Aron Depasquale, Fleet Management
- Zeljko Bulovic, Asset Manager
- Sharon Booth, Product Analyst
- George Budd, Product Manager
- Kevin Wright, Manager, Car Distribution IMX
- Fred Pietracupa, Manager Cost Analysis

Dr. Ioana-Codruţa Bilegan, currently Professor at Université de Valenciennes, France, was postdoctoral fellow and proved an invaluable resource for the project.

Ms Geneviève Hernu, M.Sc., was analyst-programmer for the project. She worked primarily to develop the simulation tool and contributed significantly to the data analysis and manipulation.

M. Julien Dubreuil undertook his master degree on *Seaport Container Terminal Operation Management*, under the supervision of Professor Crainic, concurrently with the project. This provided the project with a good deal of valuable information. M. Mohamed Ould Ebebe assisted with data analysis and manipulation during the last months of the project.

The project was managed according to the structure and committees specified by Transport Canada. The Project manager supervised the entire operation. With respect to the scientific content of the project, the project manager was assisted by a *Steering Committee* composed of Professor Crainic, Project manager, Professor Gendreau, Director of CRT, and Graham Gordon, CN. All participants in the project reported to the project manager and the Steering Committee. This committee made sure that work proceeded along the lines specified in the initial proposal and in accordance with the decisions of the *Project Management Committee* composed of Mrs. Melody Miller, Transport Canada, who replaced Mrs. Madeleine Betts, Transport Canada, during the last months of the project, Professor Crainic, CRT – Université de Montréal, and Mr. Luc Lefebvre, Ministère des Transports du Québec who represented the Canadian ITS community. Ing. Pierre Bolduc and Ing. Raynald Ledoux, Transportation Development Centre, Transport Canada, successively acted as Technical Authority for the project.

The project was managed at the Centre for Research on Transportation (CRT) with the full administrative and secretarial support of the CRT staff. Ms. Josée Vignola, Administrative Officer of CRT, supervised the administrative issues, as well as the paperwork required by the project, kept the financial accounts of the project, and prepared the required financial statements. The project was managed in accordance with the Université de Montréal's management and financial control procedures.

III. Project Timeline and General Methodology

We identified five (5) main work directions for the project. The components and schedule of each direction appear in Figure 1, which displays the final project timeline. The project started on November 1, 2004.

The first direction was made up of a single task, the review of the relevant literature and yielded the report making up Section IV.

The second, third, and fourth sets of tasks correspond to the three main work directions of the project: container-release problem (direction B), intermodal service strategy (direction C), and development of the simulation tool (direction D). Sections VI, VII, and VIII present the general description, main tasks, status, and accomplishments for each work direction, respectively. Work on all three directions required the collection, analysis, and handling of significant quantities of data. This work is detailed in Section V.

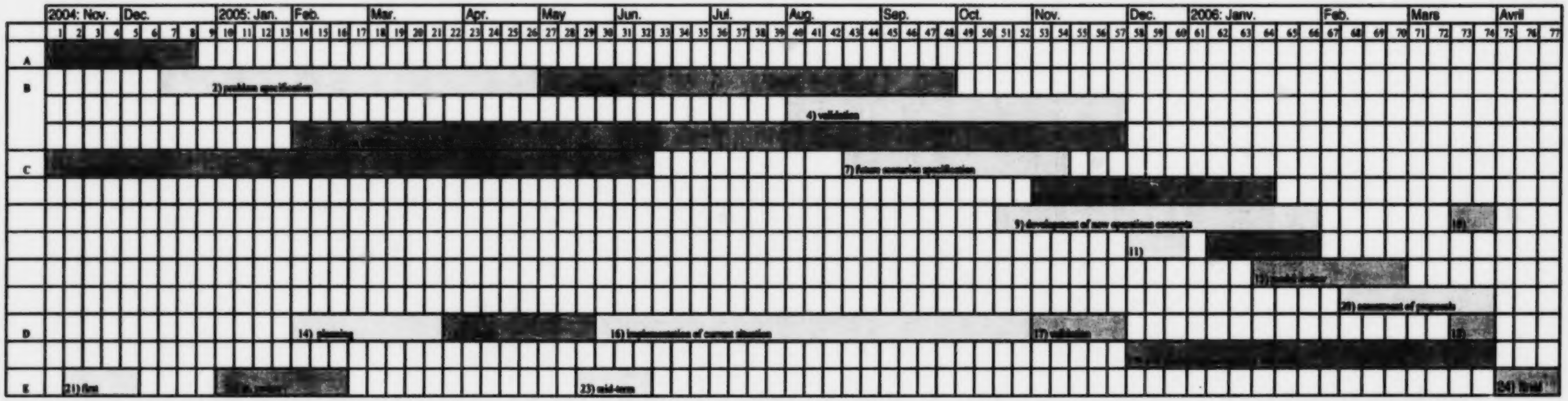
The fifth set of tasks included the production of the five reports according to the Transport Canada criteria.

The three directions of research involved the development of models or software. In order to ensure a sound and formally tractable process for these developments, we followed a general framework generally used for complex systems development. Based on general Requirements Engineering concepts and extensively used in System Engineering for complex applications (e.g., software, physical, electric, and electronic systems), this standard procedure offers a well-defined framework that applies equally well to operations research modelling activities. According to this framework, illustrated in Figure 2, the ultimate requirement – the initial definition (characterization, expression) of the main purpose of a given project – is analyzed and progressively refined, according to the requirements and constraints of each development phase. Thus, at each stage of development (planning, technical requirement specification, design, implementation, etc.), a coherent and tractable working mode is obtained, which also provides an implicit set of verification and validation methods.

Obviously, not all phases applied to our project. Installation and Maintenance, in particular, were without object in this project. Moreover, within each work direction, several methodologies were combined: operations research (model and algorithmic development), artificial intelligence (neural networks for container-release modelling), and simulation.

The project required several meetings between the university researchers and CN personnel, as well as a continuous flow of electronic exchanges, too numerous to list in this report. Several other meetings involved port and container terminal personnel in the ports on Montréal and Halifax:

- Jean-Luc Bédard, Vice-President, Operations, and Harbour Master, Port of Montreal;
- Jim Nicoll, Manager of Systems, and Ches Carter, Manager of Operations, Port of Halifax; We also participated to the Collaborative workshop (Smart Port Initiative) organized by the Halifax Port Authority;
- Murray Graves, Administrative Manager, Halterm terminal, Port of Halifax;
- Calvin Whidden, Vice-president, Ceres terminal, Port of Halifax ;
- Kevin Doherty, General Manager, Gateway terminal, Port of Montreal;
- Mr. Didier Vanal, Managing Director, CMA CGM (Canada);
- Mr. Andrew Nation, Operations Manager, CMA CGM (Canada).



A = Literature Review
B = Container Release Model
C = Intermodal Services
D = Simulator
E = Reports

1) relevant literature review
2) problem specification (includes visits to ports of Montreal, Halifax)
3) modelling
4) validation
5) data collection and analysis
6) problem description: current situation
7) future scenarios specification
8) data collection
9) development of new operations concepts
10) concept confrontation and validation with CN partners
11) requirements for integration into simulation
12) design

13) model review and refinement
14) planning and requirements
15) design
16) implementation of current situation
17) validation
18) container release model integration
19) new models integration and validation
20) assessment of proposals and of associated requirements for future development
21) first report
22) relevant literature review report
23) mid-term report
24) final-report

Figure 1. Project Timeline and Task Identification

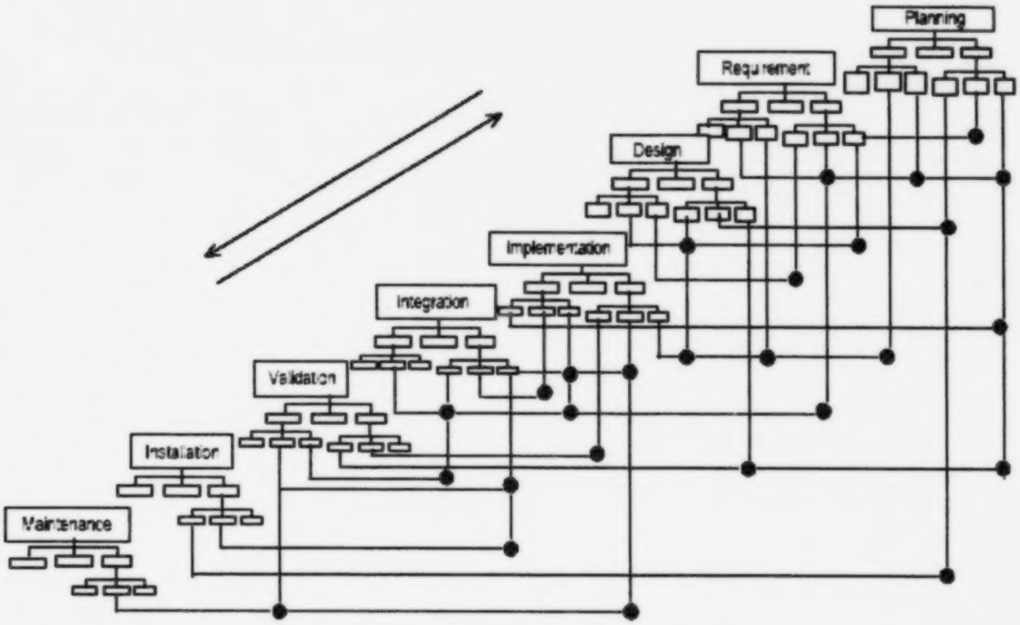


Figure 2. Abstract Software Development Model (Kececi and Modarres, 1998)

IV. Literature Review

This section starts with a general presentation of rail transportation with its main operations and physical components, a number of important planning decisions, and the corresponding main research contributions. It then examines issues and contributions focussing on rail intermodal transportation with bookings, maritime navigation and ports, as well as Intelligent Transportation Systems. The literature review relative to forecasting the release-time of containers from port to the carrier concludes the section.

IV.1. Rail Transportation

Consider the case of a railway carrier that operates networks made up of single or double-track lines that connect many large and small classification yards where rail cars are grouped and trains are formed, as well as pickup and delivery stations, junction points, etc. (Assad 1980a, Crainic 1988, Cordeau, Toth, and Vigo 1998).

The process starts when a customer issues an order for a number of empty cars or, alternatively, when freight is brought into the loading facility following a pickup operation. At the appropriate yard, rail cars are selected, inspected, and then delivered to the loading point. Once loaded, cars are moved to the origin yard (possibly the same) where they are sorted, or *classified*, and assembled into *blocks*. A block is a group of cars, with possibly different final destinations, arbitrarily considered as a single unit for handling purposes from the yard where it is made up to its destination yard where its component cars are separated. Rail companies use blocks to take advantage of some of the economies of scale related to full train loads and the handling of longer car strings in yards. The block is eventually put on a train and this signals the beginning of the journey. During the long-haul part of this journey, the train may overtake other trains or be overtaken by trains with different speeds and priorities. When the train travels on single-track lines, it may also meet trains traveling in the opposite direction. Then, the train with the lowest priority has to give way and wait on a side track for the train with the higher priority to pass by. At yards where the train stops, cars and engines are regularly inspected. Also, blocks of cars may be *transferred*, i.e., taken off one train and put on another. When a block finally arrives at destination, it is separated from the train, its cars are sorted, and those having reached their final destination are directed to the unloading station. Once empty, the cars are prepared for a new assignment, which may be either a loaded trip or an empty repositioning movement.

One source of complication in rail freight transportation is the complex nature of the yard activities: the classification of cars and the composition of trains. The modelling of yard operations as well as that of their interactions with the rest of the system are critical components of any comprehensive rail model. It is interesting to note that, traditionally, in most rail systems cars were spending most of their lifetime in yards: being loaded and unloaded, being classified, waiting for an operation to be performed or for a train to come, or simply sitting idle on a side track. Also of interest is the fact that most rail companies have separated the operations and yards dedicated to intermodal services from those used for their regular services in an attempt to cut delays, especially those associated with yard operations, and improve the quality of this time-sensitive and highly competitive service (Crainic, 2003). Research is under way to develop technology and planning and operations methods to improve the efficiency of intermodal traffic, particularly involving rail transportation (e.g., Bostel and Dejax 1998, Macharis and Bontekoning 2004, Bontekoning *et al.* 2004).

Another notion often encountered in transportation planning has to do with *schedules* and *scheduled services*. In the general sense, a *schedule* specifies timing information for each possible occurrence of a service during a given time period: departure time at the origin, arrival/departure time information at intermediary stops, and arrival time at the final destination. The schedule may also include indications on the *cut-off* time: the latest moment freight may be given to the carrier and still meet the scheduled departure of the service.

Schedules are omnipresent in passenger transportation by air, rail, bus or ship and are strictly enforced (most of the time). The case is less clear for freight transportation. On the one hand, there are no schedules in door-to-door transportation, except for cut-off times. At the other end of the spectrum, regular navigation lines usually operate according to strict schedules (high port utilization fees constitute an important incentive to follow the schedules). Much air cargo is moved on passenger planes and therefore follows strict schedules. All-cargo air services are also usually operated according to well-established schedules.

Rail freight transportation is much more varied in its approach to schedules. In Europe, for example, the railway networks are very congested and freight trains must be operated according to schedule such that freight and passenger traffic are smoothly interlaced. In many other regions of the globe, trains leave "when full", which makes schedules "indicative" only, when they exist at all. North American railways operated traditionally according to loosely-scheduled-leave-when-full policy. The creation of Intermodal divisions and other specialized services was in fact a response to this situation when attempting to offer a better service and image in particular market niches. The industry has significantly evolved in the recent years and most major railways in North America are now operating partially or totally scheduled networks (e.g., Ireland *et al.* 2004). According to our best knowledge, CN is the first railway carrier in North America, and one of the first in the world, to introduce both a booking system and a full-asset-utilization policy that runs the same train services (i.e., the same block definition and capacity) at each period.

Rail transportation systems thus appear as complex organizations that involve a great deal of human and material resources and that exhibit intricate relationships and tradeoffs among the various decisions and management policies affecting their different components. The planning process at a railway company is characterized by three main stages (Crainic 1988, Joborn 2001):

- *The strategic stage*: long-term planning, which involves the highest level of management, requires large investments, and shapes the general strategies of the system. Some of the decisions at this level of planning are concerned with physical network design and upgrading, facility location, resource acquisition, and service policy definition. In particular, for rail transportation, this stage typically includes deciding about the number of engines the company needs, determining the size of the freight car fleet, and deciding which type of freight cars the company should provide; if the railway company owns the railway track, infrastructure investments are strategic planning decisions also.

- *The tactical stage*: medium-term planning, necessary to ensure an efficient and rational allocation of existing resources to improve the performance of the whole system. At this level, data is aggregated, policies are abstracted and decisions are sensitive only to broad variations in data or system parameters. Typical tactical decisions concern the *design of the service network* and may include issues related to the determination of the routes and types of service to operate, service schedules (time table), vehicle and traffic routing, repositioning of the fleet for use in the next planning period. In a scheduled railway, the time table defines the days of operation, routes, stops, arrival and departure times, and transportation capacities of the trains. The plan and schedule of operations also defines the blocks that are built at the yards and which connections between trains are possible. Taking into account the handling times at classification yards, it usually defines all the possibilities for transporting a car from an origin to a destination. The plan and time table are usually revised a few times a year.

- *The operational stage*: short-term planning, which is performed by local management in a highly dynamic environment where the time factor and detailed representations of vehicles, facilities and activities are essential. Important operational decisions concern the maintenance activities, the scheduling of crews, the dynamic allocation of scarce terminal resources, the allocation and repositioning and of assets, engines, freight cars, crews, etc., are determined at this stage.

This classification highlights how data flows among decision-making levels and how policy guidelines are set. The vertical or hierarchical relation specifies how the data flows and how the policy guidelines are set: the strategic level sets the general policies and guidelines for the decisions taken at the tactical level which, in turn, determines goals and limits for the operational decision level which regulates the transportation system; the data flows follow the reverse route, each level of planning supplying information essential for the decision making process at a higher level. This hierarchical relationship emphasizes the differences in scope, data, and complexity among the various planning issues, prevents the formulation of a unique model for the planning of freight transportation systems, and calls for different model formulations that address specific problems at particular levels of decision making. Figure 3 illustrates the decision-making hierarchy in planning rail operations, as seen by Assad (1980b), Crainic (1982), and Kwon (1994).

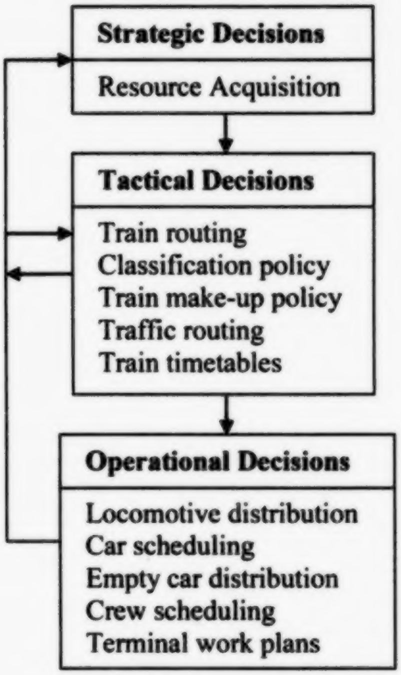

Strategic Decisions
Resource Acquisition
Tactical Decisions
Train routing
Classification policy
Train make-up policy
Traffic routing
Train timetables
Operational Decisions
Locomotive distribution
Car scheduling
Empty car distribution
Crew scheduling
Terminal work plans

Figure 3. Hierarchical Decisions in Rail Operations

At each planning level, particularly at the tactical planning level, a second type of relation, which may be called "horizontal", can be observed amongst the various decisions. Indeed, all tactical level planning problems and the policies established to deal with them have network-wide impacts and are strongly and complexly interrelated in both space-time dimensions and economic aspects (Crainic 1988). For example, if in a rail system, certain new direct services (bypassing intermediate yards) are introduced, some possible consequences are: a decrease in congestion in some yards, lower delays in those yards, lower transportation times (at least theoretically) for some traffic, and higher transportation costs. On the other hand, a higher number of direct trains might imply higher congestion on the lines, which implies higher delays for lower priority services, generating higher transportation times for some traffic, etc.

The number of complex problems in the field of rail transportation that can be modeled and solved using mathematical optimization techniques is very large. However, the related literature has experienced a slow growth and, until recently, most contributions were dealing with simplified models or small instances failing to incorporate the characteristics of real-life applications (Cordeau *et al.* 1998). Indeed, the literature surveys of Assad (1980a, 1981) and Haghani (1987) suggested that, at that time, carriers used more simulation-based tools than optimization methodologies. Nowadays, although simulation-based approaches are still widely used to evaluate and compare different scenarios, one witnesses a sustained development of optimization methods capable of producing high-quality solutions to complex problems within short computing times.

A number of reviews of models and methods for planning rail operations may be found in the literature. Assad (1980a, 1981), Haghani (1987), and Crainic (1988) review early contributions to the field. Cordeau, Toth, and Vigo (1998) survey optimization models for the most commonly studied rail transportation problem, i.e. routing and scheduling problems, with regard to both freight and passenger transportation. The review addresses, in particular, methodologies for freight transportation and railcar fleet management. Kraft (1998) presents a rich rail-related literature review classified by function (network operations, equipment distribution, road operations, yard operations) and planning level. Crainic (2000, 2003) and Crainic and Laporte (1997) review models that address strategic, tactic, and operational issues for railways in the larger context of freight transportation. Crainic and Kim (2005) perform the same function for intermodal transportation. In the next two subsections we present a very brief survey of methodologies for service network design (tactical planning) and fleet management (operations planning).

Service Network Design

The term *service network design* covers tactical planning issues: on what routes to provide service, what type of service (mode) to use, how often to offer service on each route and according to what schedule, how to route the loads through the physical and service networks, how to distribute the work among the terminals of the system, the goal of the planning process being a transportation plan that ensures customer satisfaction through efficient and profitable operations and utilization of resources. Service network design formulations are used to build a transportation (or load) plan to ensure that the system operates efficiently, answers demand, and ensures the profitability of the firm.

In general, customers not only expect low tariffs, but also require a high quality service, mostly in terms of speed, flexibility, and reliability. The significant increase in the market share achieved by motor carriers is due to a large extent to this phenomenon. Consequently, one of the major objectives of tactical planning, particularly for rail and intermodal transportation, is to achieve the best trade-off between operating costs and firm profitability, and service performance measured, in most cases, by delays incurred by freight and vehicles or by the respect of predefined performance targets.

Service network design problems thus address two types of major decisions. The first is to determine the service network, that is, to select the routes – origin and destination terminals, physical route and intermediate stops – on which services will be offered and the characteristics of each service: mode, frequency or schedule, etc. The second major type of decision is to determine the distribution of traffic, that is, the itineraries (routes) used to move the flow of each demand: services used, terminals passed through, operations performed in these terminals, etc. The service network specifies the movements through space and time of the vehicles and convoys of the various modes considered. Blocking decisions and operating rules specifying how cargo and vehicles are to be sorted and consolidated at particular terminals are also part of the process of designing the service network. The itineraries used to move freight from origins to destinations determine the flows on the services and through the terminals of the service network.

These problems and decisions have network-wide impacts and are strongly and complexly interconnected both in their economic aspects and the space-time dimensions of the associated operations. Therefore decisions should be made globally, network-wide, in an integrated manner (Crainic and Roy 1988). More formally, main decisions made at the tactical level concern the following issues:

1) *Service selection*: The routes – origin and destination terminals, physical route and intermediate stops – on which services will be offered and the characteristics of each service. Frequency or scheduling decisions are part of this process.
2) *Traffic distribution*: The itineraries (routes) used to move the flow of each demand: services used, terminals passed through, operations performed in these terminals, etc.
3) *Terminal policies*: General rules that specify for each terminal the consolidation activities to perform. For rail applications, these rules would specify, for example, the blocks into which cars should be classified (the *blocking* policies), as well as the trains that are to be formed and the blocks that should be put on each train (the *make up* rules). An efficient allocation of work among terminals is an important policy objective.
4) General *empty balancing* strategies, indicating how to reposition empty vehicles to meet the forecast needs of the next planning period.

Several efforts have been directed toward the formulation of tactical models. See the reviews of Assad (1980a), Crainic (1988, 2000, 2003), Delorme, Roy, and Rousseau (1988), Cordeau, Toth, and Vigo (1998), Crainic and Laporte (1997), Crainic and Kim (2005). Network models, which take advantage of the structure of the system and integrate policies affecting several terminal and line operations, are the most widely developed. Simulation models have been proposed and used by transportation firms to evaluate scenarios and select policies. Network optimization formulations, on the other hand, may efficiently generate, evaluate, and select integrated network-wide operating strategies, transportation plans, and schedules. These models are discussed in detail in Crainic (2003). Most service network design and related issues yield fixed cost, capacitated, multicommodity network design formulations. These formulations may be static or dynamic but, up to now, have been generally deterministic. The former assume that demand does not vary during the planning period that is considered. The time dimension of the service network is then implicitly considered through the definition of services and inter-service operations at terminals. Time-dependent formulations include an explicit representation of movements in time and usually target the planning of schedules to support decisions related to when services depart.

Fleet Management

Most of the tasks at the operational level are performed in a highly dynamic environment where the *time* factor plays an important role not only in the reaction time available to a changing environment, e.g., new customer requests or service perturbations, but also in the representation of decision outcomes and the selection of the most appropriate one.

Many models traditionally used in transportation planning use known static data as their input. Tactical planning formulations, for example, consider aggregated forecast demand data as "known". However, the real world in which these models are used is in a constant state of change and solutions cannot always be implemented as planned. If traffic is slower than predicted, vehicles may arrive late at customers' locations or at the terminal. Forecasted customer requests for empty containers or railcars may not materialize while unexpected demands may have to be satisfied. The planned supplies of empty vehicles at yards may thus be unsettled and additional empty movements may have to be performed. Consequently, the dynamic aspect of operations is further compounded by the *stochastic* characteristic inherent to the system, that is, by the set of uncertainties that are characteristic of real-life management and operations. Increasingly, these characteristics are reflected in the models and methods aimed at operational planning and management issues, as illustrated in Crainic (2003).

A significant step forward in modelling capabilities was achieved with the explicit consideration of the time perspective. A space-time network represents the various paths that vehicles may travel to reach their proper destination at a specified time (including, eventually, holding decisions at terminals). The resulting formulation takes the form of a deterministic time-dependent transshipment network model, where flows are optimized such that either the total cost is minimized, or the profitability of the system is maximized. Starting with the pioneering contributions of White (1968) and White and Bomberault (1969) for rail car distribution, and of White (1972) for container allocation, many models that aimed for the distribution of empty vehicles, took the form of a dynamic transshipment network optimization problems solved by using linear programming and network flow algorithms (e.g., Ermol'ev, Krivets, and Petukhov 1976 and Florez 1986). More details on these early approaches may be found in Dejax and Crainic (1987).

The explicit consideration of uncertainties in empty vehicle distribution models constitutes another significant methodological contribution. The first comprehensive effort in this direction was made by Jordan and Turnquist (1983) for rail. The formulation aims to maximize profits and integrates revenues from performing the service as well as various costs from moving cars between yards, holding cars at yards, or from not filling orders due to stockouts. The model structure is again a multicommodity, dynamic network, explicitly integrating the stochasticity of supply, demand, and travel times. Many developments followed, most of them applied to rail and motor carrier issues (Crainic 2003, Cordeau, Toth, and Vigo 1998, Powell 2003; Powell and Topaloglu 2003, 2005, Powell, Bouzaïene-Ayari, and Simaõ 2005).

Fewer efforts were dedicated to container fleet management issues. Crainic, Gendreau, and Dejax (1993) proposed a series of models for the allocation and management of a heterogeneous fleet of containers where loaded movements are exogenously accepted: single and multicommodity deterministic formulations and a two-stage, restricted recourse single commodity, stochastic model. Cheung and Chen (1998) presented a two-stage stochastic model for the single-commodity dynamic container allocation problem for liner operators (regular ocean navigation lines). Powell and Carvalho (1998) address the problem of the combined optimization of containers and flatcars for rail intermodal operations.

IV.2. Intermodal Systems with Reservations

A *scheduled-with-bookings* system operates regular scheduled services and customers have to book space in advance in order to use the service. Each class of customers or destination has a certain space allocated on each train and potential clients are required to call in advance and reserve the space they need. This new approach to operating intermodal rail services brings advantages for the carrier, in terms of operating costs and asset utilization, as well as for customers in terms of increased reliability, regular and predictable service and, eventually, better price.

Scheduled-with-bookings systems require the same type of planning methods for their service and operations as regular ones (see Section IV.1). Yet, their particular characteristics lead to significant differences that require revisiting models, methods, and practices. The field is very new and, consequently, very little literature exists, however.

It is noteworthy that passenger intermodal transportation cannot be an example for freight intermodal transportation due to the many conceptual differences between the two (Bontekoning *et al.* 2004). Thus, for example, passengers "move" themselves between two modes of transportation, whereas goods need to be moved, using specialized equipment, either individually or in batches. Moreover, arrival and loading processes for containers have "industrial" characteristics, in terms of volume, dimensions, timing, and handling requirements that are very different from the requirements of human passengers. Consequently, we do not review the literature relative to systems with reservations in passenger rail transportation.

We are aware of only two lines of research directly relevant to rail operations with bookings. Both address operational issues only.

Kraft (1998) was among the first to address issues related to the operation of rail systems with reservations. The goal of the research was determining the optimal routing for a set of shipments assuming a fixed capacity network of blocks and trains, as well as the service commitments that should be offered to customers based on this same set of operating and capacity constraints. Kraft ended up by developing a method for managing day to day railroad network operations through a reservations-based, capacity-constrained car scheduling process. Two models were proposed, a train segment pricing model and a dynamic car routing model.

The train segment pricing model aimed to determine the service offer to make on customers requests, in order to give decision support to the reservation center of the company. It is a profit-maximizing revenue-management model that determines service offers to maximize expected profit, taking into account costs, the associated revenue, the probability the demand will materialize, and the probability the customer will accept the offer, as well as the availability of equipment and the line haul capacity to meet service commitments and the requirement that, once a shipment (booking request) is accepted, it must be managed consistently with its delivery commitment. It is a bid-price-based approach that should allow a carrier to develop in real-time achievable (i.e., no service failure) and market-sensitive quotations (service offers) of delivery time for new shipments calling in. The implementation made use of forecasted demands for 7-10 days into the future.

The second model was developed to support the needs of classification yards by suggesting blocks and trains in which each car should ride. It is a deterministic, costminimizing, multi-commodity network flow model to be applied to the "initial inventory" of shipments currently on line only. The problem is to deliver all shipments within their committed delivery times at minimum cost. The current position and destination of each car are known. Each car has an agreed-upon delivery time window, with a penalty cost for early or late arrival. Penalty costs represent actual payments or rate reductions, an estimate of loss of goodwill, loss of future revenues or profits, or some combination of these. This model dynamically reroutes shipments to take advantage of all available train capacity in the network, while still meeting the committed delivery times on priority shipments.

Both procedures have been integrated into a rolling-horizon simulation model. Simulation results indicated that up to 10% improvement in the railway operating ratio could be achievable though implementation of this shipment management strategy. Kraft and Guignard-Spielberg (1993) presented a terminal operations model compatible with the dynamic car routing proposed in Kraft (1998). A study of demand variability and its impact on service reliability is presented in (Kraft 1995).

The other study related to scheduled-with-bookings freight rail transportation that we know of addresses the issue of optimizing the empty freight car distribution in scheduled railways (Joborn 2001, Holmberg, Joborn, and Lundgren 1998, Joborn *et al.* 2004). The Swedish Railways constituted the focus of the studies, but the results could be applied to other railway companies as well. Because the research was related to a commercial application, all aspects of the new distribution system could not be disclosed. However, detailed description of the optimization aspects of the system, as well as information about their setting and usage were given, for a good presentation of the optimization approach.

Schematically, the Swedish railway system with schedules and reservations operates as follows. All planned transports of cars are booked in a reservation system for the freight trains. As soon as the demand for loaded transports is known, the corresponding loaded freight cars are booked onto the trains. The loaded car movements for the next 24 hours have to be booked before noon each working day, and after noon the empty freight cars are booked. The remaining capacity of the train after the loaded cars have been booked can be used for empty cars. When planning the empty car distribution, the loaded transports cannot be altered, which means that the remaining capacity of the trains for empty freight car transports can be considered as fixed. If a train is over-booked, so that its maximum weight or length is exceeded, the booking of some of the cars are changed to other departures on the same day. If some trains are still over-booked after these changes, some cars have to be left behind. As a consequence, the car that was left behind will not reach the destination at the time expected. In general, empty cars are more often left behind than loaded ones.

Reservations should make sure that there is enough transportation capacity for the planned transports for both loaded and empty freight cars, all the way from origin to destination. Together with the time table, reservations also "guarantee" that planned empty movements can be performed. Consequently, the empty car distribution problem becomes deterministic.

The distribution planning system developed and implemented at Swedish Railway is used in an on-line setting. The system continuously calculates the availability of freight cars for each demand for empty cars reported. It also makes reservations in freight trains to secure delivery of empty freight cars. The model considers several operational issues that may be critical to performance: capacity restrictions on trains, gross (instead of net) values of supply and demand, extended planning period (over several days), and substitution possibilities between freight car types. Furthermore, the model includes representation of detailed requirements of the transport structure of the railway system, such as different sizes of freight car types, weight and length restrictions of trains, forward-sending delay, possible train connections, and shunting movements. The optimization model and algorithm implemented at Swedish Railways is successfully used for continuous, operational distribution planning. However, one interesting conclusion the author reached was to never neglect the importance and difficulties of getting correct input data. The on-line update of input data and short term predictions about the future supply of freight cars contributed in particular to the complexity of the input data handling.

We did not find any simulation model directly aimed at general rail intermodal systems with reservations. The work of Gambardella *et al.* (2002) addresses issues for a very particular case but it illustrates the possibilities offered by simulation in the larger context.

Gambardella *et al.* (2002) developed a simulation model for the flow of Intermodal Terminal Units (ITU) among inland intermodal terminals. The work was part of the PLATFORM project, funded by the Directorate General VII of the European Community. One of the objectives of the PLATFORM project was the implementation of a simulation environment for the assessment of impacts produced by the adoption of various technologies and management policies aimed at improving terminal performance.

A brief description of the simulated environment is given in the following. An intermodal terminal can be regarded as a node in a network that models the connectivity of the origins and destinations in the supply chain. The intermodal terminals are interconnected by rail corridors. Each terminal serves a user area via a road network. The terminal is modeled as a set of platforms, which are served by a number of gantry cranes and front lifters. Given the schedule of train connections among the terminals, an agent-based system, the Intermodal Transport Planner (ITP) books ITUs on trains and assigns trucks to deliver them to the source terminal and to pick them up in the destination terminal. The terminal and rail corridor simulation software has been implemented as a discrete-event simulation model, using MODSIM III as development tool.

The PLATFORM architecture consists of two subsystems: 1) the intermodal transport planner (ITP) that manages the planning of the whole intermodal transport chain from origin to destination for an ITU; 2) the simulation system (composed by the road simulation, rail simulation, and terminal simulation modules) that models and simulates the ITU inter-terminal transport process in quite great detail. The ITP plans the intermodal transport task (ITT) for an ITU by using:

- Intermodal Planning and Execution Units (IPnEU) – for planning the whole ITT of an ITU. They split the ITT into its three main parts, the initial and final leg on the road and the main leg by train. They contact the specialized agents for planning, booking, and reservation of these parts.
- Forwarding Agents – for planning and booking the ITT of the ITU by truck. These agents are responsible for the planning of delivery ITUs to and their pick-up from

terminals. Each forwarder is modeled by such a forwarding agent. A broker agent coordinates the planning of the forwarding agents of the area around each terminal

- Booking Agent – for booking the ITT of the ITU by train. This agent checks for availability of places on scheduled trains, checking which bookings are possible. The booking agent then chooses the best one and makes the reservation.

Many other simulation studies and systems have been developed for intermodal freight transportation and for intermodal container terminals in general, but scheduled-with-bookings systems were not the focus of these: Kozan (1997), Gambardella *et al.* (1998, 2001), Sgouridis and Angelides (2002), Kia *et al.* (2002), Martinez *et al.* (2004), to cite just a few.

We conclude this section with two remarks motivated by the literature review just presented. There is no contribution that targets scheduled-with-bookings intermodal services at the scale of a major North American rail carrier that covers Canada and part of the United States. There are very few contributions that address issues relevant for scheduled-with-bookings systems. In particular, we found no published reference to any model or method for the design of such services. An important investment of resources and creative efforts has to be deployed in order to propose models and functional prototypes that could be tested, improved, integrated and put to work within the CN Intermodal service network.

IV.3. Ships and Ports

The previous sections targeted the *supply side*, the rail freight transport industry, of the intermodal freight transport process. In this section, we focus on the critical elements constituting the *demand side*: shipping companies, as they are transporting the containers to ports, and the ports themselves. We briefly describe their operations and identify possible sources for the delays these intermediaries introduce in the intermodal transportation chain.

Steamship lines

Ship operations and planning issues are different from those experienced by other transportation modes because ships are confronted to different conditions (Ronen 1983, 2002, Christiansen *et al.* 2005). Note in particular that ships pay port fees; the draft of a ship is a function of the weight of the load and affects ship-port compatibility; ships operate mostly in international trade (crossing multiple jurisdictions); ships can be diverted at sea; ship voyages span days or weeks and their time in port may cover several port operating time windows. Most of these characteristics are different from those of land-based carriers. In some aspects, however, aircrafts are similar to ships. For instance, both ships and aircrafts experience high uncertainty in their operations because of their higher dependence on weather conditions and technology, and because they usually straddle multiple jurisdictions.

Ronen (1983, 1993) published the first surveys of models and methods for ship routing and scheduling. The increasing interest in maritime transportation is reflected in some recent publications. A few years ago, *Transportation Science* devoted an issue to maritime transport (Psaraftis 1999). In a recently published book, Perakis (2002) gives an overview of models for a few selected problems in fleet operations and deployment. Chajakis (2000) presents six small typical case studies in marine petroleum logistics, where the use of operations research methods produced benefits for the marine carrier, the shipper, or both.

An even more recent review, presented by Christiansen *et al.* (2005; see also Christiansen, Fagerholt, and Ronen 2004), has almost 60 references published during the last decade, in the area of ship routing and scheduling. The reviewed papers were divided into strategic ship planning issues (such as design of optimal fleets and maritime supply chains), tactical/operational ship scheduling within industrial shipping and tramp shipping (e.g., optimal assignment of cargoes to ships and, hence, ship schedules), liner shipping (such as network design and fleet deployment), and other shipping issues related to routing and scheduling. Ronen's review in 1993 included only about 30 references for the former decade, in the same field. This illustrates the increasing interest of researchers in the topic.

Concerning the potential delays characterizing this particular part of the intermodal transportation chain, we identified a few factors that may cause deviations from the initial time schedule. Prior to ships' arrival in port, weather appears to be an important cause of delays. With respect to Canadian ports, mainline shipping routes cross some of the stormiest seas – the North Atlantic and the North Pacific oceans. Delays in winter in particular are measured in hours and occasionally in days. In river ports, such as Montreal, periods of low water may force ships to reduce speed. In a recent study (Transport Canada 2003), it was found that 53% of all sampled voyages on the St. Lawrence River in winter involved delays.

Ports and container terminals

Historically, ports have been important nodes of interchange between transportation modes. As supply chains become ever more extensive and complex, the role of ports and container port terminals is assuming an ever greater significance. Container traffic is being concentrated at a few hubs, and it is at these centres that some of the most important potential delays and congestion, with resulting added costs, can occur.

The main function of a container port terminal is to provide transfer facilities for containers between sea vessels and land transportation modes, trucks and rail in particular. It is a highly complex system that involves numerous stakeholders (e.g.,, port and terminal authorities, shipping companies, railroads, motor carriers, brokers, shippers, forwarders, and regulatory agencies), pieces of equipment, operations, and container handling steps. Beyond the number and variety of stakeholders and activities, the complexity is mainly due to the complex physical and informational interactions among stakeholders, on the one hand, and the different planning and operational processes taking place at the terminal, on the other hand (Crainic and Kim 2005, Steenken, Voß, and Stahlbock 2004, Vis and Koster 2003). The assignment of resources to tasks and the scheduling of these tasks are thus among the major container port terminal planning issues.

Three main areas make up a container terminal. The *sea-side* area encompasses the quays where ships berth and the quay cranes that provide the loading and unloading of containers into and from ships. The *land-side* area provides the interface with the land transportation system (the so-called hinterland of the port) and encompasses the truck and train receiving gates, the areas where rail cars are loaded and unloaded, and the associated equipment. Trucks are generally loaded and unloaded directly in the *yard area*. This third area is dedicated for the most part to stacking loaded and empty containers for import and export (in some terminals, facilities are also provided for the loading and loading of containers). Various types of yard cranes are associated with this area. So-called *transporters*, primarily yard trucks or automated vehicles, move containers between the three areas. Figure 4 (Park 2003) illustrates part of a container port terminal. One ship and three quay cranes are displayed in the sea-side area, while only trucking is shown in the land-side area. Twelve container stacks are displayed in the yard area, as well as one type of yard crane used to transfer containers between yard transporters and outside trucks and stacks, as well as to change the position of containers in the yard as required.

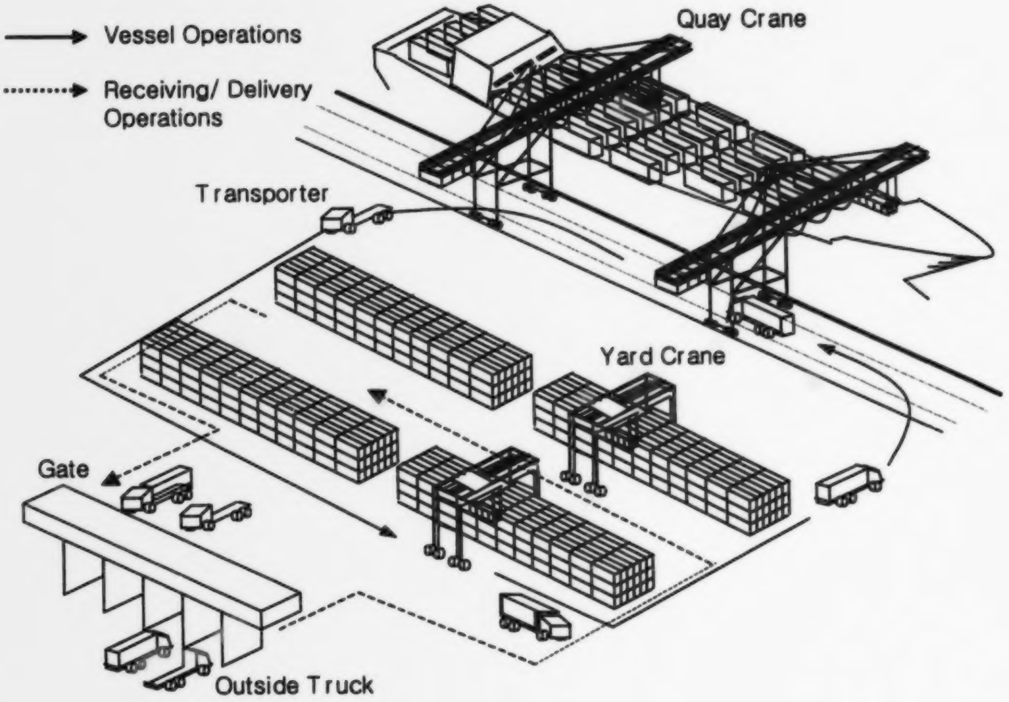


Figure 4. Container Terminal with an Indirect Transfer System (Park 2003)

Crainic and Kim (2005) developed a comprehensive literature survey related to operations and planning in container port terminals. Various issues like berth scheduling, quay crane scheduling, stowage planning and sequencing, storage space planning, allocation and dispatching of yard cranes and prime movers are addressed in their survey, and the corresponding analytic models for operational planning and control are described. Another comprehensive and recent classification and literature review of container terminal operations is given by Steenken, Voß, and Stahlbock (2004). They describe each of the main logistics processes and present the corresponding literature, including the optimization methods for each type of problem. The paper by Vis and de Koster (2003) gives an extensive survey of models and methods used for the decision and optimization problems that arise at container terminals.

Three main types of handling operations are performed in a container terminal: 1) ship operations associated with berthing, loading, and unloading container ships (see Figure 5), 2) receiving/delivery operations for outside trucks and trains, and 3) container handling and storage operations in the yard. When a ship arrives at the container port terminal, it is assigned a berth and a number of quay cranes.

When a ship arrives at the port, it has to be allocated a berth. Berth space is a very important resource in a container terminal (construction costs to increase capacity are very high, even when space for growth exists) and *berth scheduling* determines the berthing time and position of a container ship at a given quay. Lack of direct access to berths is a common source of delays. This is particularly evident in common-user berths. Major carriers will normally obtain priority berthing rights (which force smaller carriers to wait for an available berth, or may be compelled to vacate a berth if a major carrier requires access). This problem is particularly acute for feeder and coastal services. The problem is normally less acute in dedicated terminals, where the carriers arrange their schedules to fit available vacant slots.

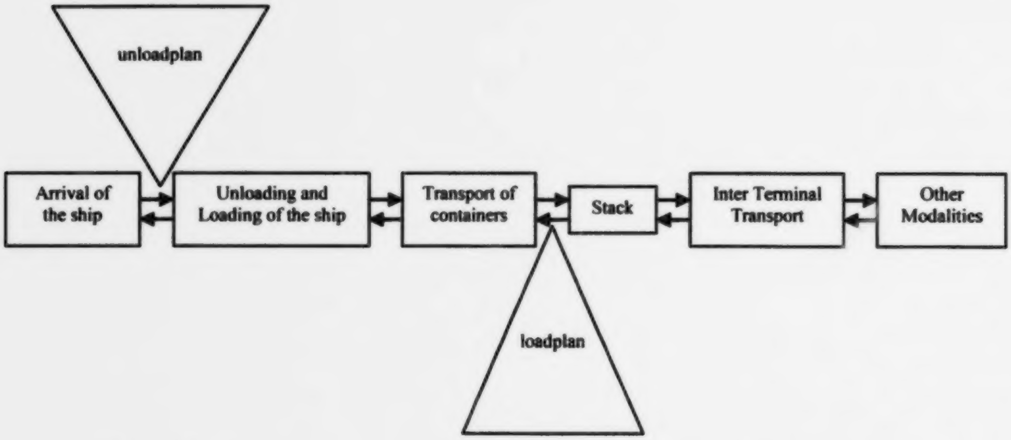


Figure 5. Processes at a container terminal (Vis and de Koster, 2003)

Once the ship berthed, the import containers have to be taken off the ship. This is done by quay cranes, which take the containers off the ship's hold or the deck. *Quay-crane allocation* is the process of determining the vessel that each quay crane will serve and the time during which the quay crane will serve the assigned vessel. The berth schedule and the quay crane allocation are interrelated because the number of quay cranes assigned to a vessel impacts the berthing duration of the ship. Despite this interrelationship, most studies treated the two issues separately to avoid the complexity of the integrated problem. The study by Park and Kim (2003) is an exception.

Similarly to the phase prior to the ships' arrival in port, at quayside also weather can be a factor of delays in many ports (Merkle 2004). Fog makes it unsafe to operate the gantry cranes, and high winds also limit the safe margins of operation from the ship to shore movement of containers (and vice versa). Snow storms are a factor disrupting handling in Canadian ports.

Stowage sequencing determines the sequence of unloading and loading containers, as well as the precise position each container being loaded into the ship is to be placed. During the unloading operation, a quay crane transfers containers from the ship to transporters. Then, the transporter delivers the import (unloading) container to a yard crane that picks it up and stacks it into a given position in the yard. This sequence of operations is called *indirect transfer*. Some terminals use a *direct transfer* system where the equipment used to move containers between the quay and the yard will also stack them. For export (loading) operation, the process is carried out in the opposite direction.

On the land-side, the *receiving* and *delivery* operations provide the interface between the container terminal activities and the external movements. A receiving operation starts when containers arrive at the gate of the terminal carried by one or several outside trucks or a train. Containers are inspected at the gate to check for damages (to the container not its content) and whether all documents are in order. Also at the gate, information regarding where the container is to be stored is provided to the truck driver. When the outside truck arrives at the indicated transfer point, a yard crane lifts a container from the truck and stacks it according to the plan. When containers arrive by rail, the rail cars are brought in the rail area where containers and documents are examined. Containers are then transferred by a gantry crane to a transporter, which delivers them to the yard and stacks them. In the case of a delivery operation, the yard equipment delivers a container onto an outside truck, which leaves the port, or onto a transporter which delivers the container to the rail area and loads it onto the designated car.

Hours of operation at the terminal gate may have an impact on delays (e.g., truckers may congest the entrance at opening time, or prior to gate closing). Along the rail distribution network, late arrival of trains (due to weather or other factors) may also have an important impact on delays. A particular problem for some US railroads are the performance guarantees given to priority customers. Customers such as UPS have delivery guarantees, the penalties for which are severe. Thus the railroads clear the tracks as much as possible for such trains, causing others along the line to be shunted aside, producing delays in scheduled services.

The sea and land-side operations interact with the yard container handling and storage operation through the information on where the containers are or must be stacked within the yard. How containers are stored in the yard is one of the important factors that affect the turn-around time of ships and land vehicles. The *space-allocation problem* is concerned with determining storage locations for containers either individually or as a group. Yard storage space is pre-assigned to containers of each ship arriving in the near future to maximize the productivity of the loading and unloading operations.

A container yard consists of blocks of containers, which are separated by aisles for transporters as shown in Figure 4. A block consists of 25 to 35 yard bays, and a yard bay has 6 to 10 stacks of containers. Container handling and storage operations include the management and handling of containers while they are in storage in the yard and thus occur between the receiving and delivery operations and the ship operations. Container-handling equipment (cranes or straddle carriers that can both transport containers and store them in the stack) performs the placement of containers into storage and their retrieval when needed. Yard cranes move along blocks of containers to yard bays to perform these operations. Planning these operations is part of the *equipment-assignment* process, which allocates tasks to container-handling equipment. Based on the quay-crane schedule, one or two yard cranes are assigned to each quay crane for loading and unloading. The remaining yard cranes are allocated to receiving and delivery operations. Terminal operators aim to assign and operate yard cranes in such a way that inefficient moves and interferences among yard cranes are minimized.

Related to operations within the terminal, delays due to mechanical failures of equipment are to be expected. However, congestion upon arrival of large vessels or just prior to their departure, large traffic volumes that have to be positioned prior to loading, and seasonal increases in traffic volumes, constitute the most important sources of delays. Labour issues may also significantly impact operations (e.g., the US West Coast dock workers' strike halted 60% of all US imports for several weeks).

With respect to the system-wide delays, *congestion* is the biggest problem that can occur anywhere along the chain and may have system-wide consequences. Other than the instances referred to above, the most common source of congestion is traffic growth. Seasonal or systemic increases in traffic may impact severely throughout the chain. The ports of the American West Coast are experiencing such conditions today with the growth of trade with China. According to the Los Angeles-Long Beach Marine Exchange, which monitors vessel traffic at both ports, the turnaround times for ships calling at Los Angeles or Long Beach is currently averaging four to five days, with some taking between eight and 10 days (CalTrade Report 2004). While congestion has been most evident on the West Coast, it spread to the East Coast ports too in 2004 (Armbruster 2004). Similar issues are observed in Canadian West Coast container terminals.

A more localized congestion may occur where a shipping line adds a new port of call to its routes. For a major carrier this may involve the immediate addition of up to 1000 box transfers per call and may overwhelm the capacity of the terminal, at least in the short run. Conversely, a berth may find itself with excess capacity when a shipping line quits the port. Shipping lines add and drop ports with considerable frequency.

Security has been identified as a source of localized delays and disruptions. Customs have traditionally searched approximately 2% of containers. Customs procedures require containers to be removed from onward distribution until cleared. On certain routes and during certain periods containers were being used to transport illegal immigrants and their discovery would force searches of other boxes aboard the ship. This would inevitably delay their discharge by a considerable amount of time.

As a consequence of the *post 9/11 security* requirements, the US has imposed strict security regulations on containers, in recognition of the extreme importance of container transport to the economy and the difficulty of examining goods that are housed in enclosed boxes. Of the many regulations introduced, most important in terms of potential delays are: the Container Security Initiative (CSI), the 24 Hour Rule, and the US Trade Act of 2002 (UNCTAD 2004). These three new regulations are briefly explained below:

- **CSI:** This initiative is directed at identifying potentially high-risk containers and screening them before they are put aboard ships; US Customs personnel are placed in certain foreign ports to screen containers as they wait in the container yard prior to their lifting aboard ship.
- **The 24 Hour Rule:** US customs now demand detailed manifest information for US-bound traffic at least 24 hours before loading at a foreign port. The information must

provide a detailed list of all the contents, unlike before when 'FAK' (freight-all-kinds) could suffice, and must be provided by the carrier, not by the shipper.

- **US Trade Act of 2002:** This act extended the requirements for mandatory reporting of container contents to US export cargo as well as imports.

It is still too early to evaluate the full impact of these initiatives. In terms of delays, the 24 Hour Rule and the US Trade Act are likely to be the most important. Because carriers have the burden of reporting manifests, they are demanding ever earlier closing times. Before, it was not uncommon for them to allow 'late gate' practices, where preferred customers could deliver a container to the terminal up to 6 hours before vessel departure. There is growing evidence that shippers are building buffer periods into their logistics operations. The problem is boxes that have been delivered to a terminal and have not yet been cleared to depart; this represents an additional burden on storage and a potential source of congestion.

Canada adopted a 24 hour rule for imports that came into effect in April 2004. In December 2002 the International Maritime Organization (IMO) agreed to introduce a new security regulation, the International Ship and Port Security (ISPS) code, which came into effect in July 2004. Contracting nations are to determine and set security levels for vessels flying their flag, issuing International Ship Security Certificates (ISSC), and carrying out port facility security plans. Automatic information for ships has been included in the provisions. Ships must provide their ISPS certificates to port authorities where they make calls.

The major impact reported in the trade literature is one of costs involved, the burdens being placed on port authorities and shipping lines to comply with the new regulations. There is also a great deal of discussion about delays. Inspection of containers in ports is the most frequently reported item. An overall average in the US suggests that 10% of containers are now inspected, a total in excess of 900,000 TEUs (Twenty-Foot Equivalent Units). In particular cases, the figures are higher, with in the port of Miami 25% of containers being screened by new X-ray machines, and 10% being inspected physically. Delays and congestion are reported (Containerisation International 2003). This is to be expected to occur in Canadian seaport container terminals that have to comply with the requirement to install X-ray machines close to the ship, so that containers are screened before being stacked in the yard.

In conclusion, the main types of time delays that could be identified and which should be taken into account in the evaluation of the overall time of cargo in port, before its departure by another mode of transport, are given in the following, not exhaustive, list:

- Ship's waiting time to be berthed;
- Ship's waiting time to start discharging operations;
- Time necessary for unloading of containers (which depends on the handling rate of discharge operations);
- Time from quay to storage;
- Time in storage;

- Time spent in transferring cargo from storage to another mode of transport (including loading time);
- Time spent in carrying out logistics activities required by customers that add value;
- Time for goods to be cleared (if such is to be done at port level);
- Time spent by cargo awaiting departure of the next mode of transport (road or rail).

Research on automation of equipment and activities in ports is being pursued actively and real-time decision problems are of increasingly interest. Pointers toward related literature and details about these issues can be also found in the book chapter by Crainic and Kim (2005). Simulation models have been used frequently to evaluate policies and operational rules of conventional, as well as of automated container terminals, yet there is no study of global optimization of operations within an integrated framework of the whole port platform.

Related issues: agile and lean ports

The ability to respond rapidly to markets that are driven by sudden changes in customer demand is called *agility*. This implies flexibility and enables business to grow in competitive markets of continuous and unanticipated change (Yusuf *et al.* 1999). Interesting and relevant for intermodal transportation is the study of agile ports given in the paper of Marlow and Paixão Casaca (2003).

The development of agile ports requires the efficient application of knowledge, i.e. change proficiency, since the environment in which ports carry out their operations is governed by a knowledge-based economy. In this sense, this notion is related to the utilization of Intelligent Transportation Systems methods and technology. Port efficiency is an important determinant of ocean freight rates. Agile ports are capable of exploiting unanticipated customer opportunities which allows them to become providers of integrated transport solutions. The main problems that must be faced in such an environment are the volatile demands caused by market uncertainty, time sensitive customers, random or seasonal demand of multimodal processes, operational and logistical constraints, changes in trade flows, etc.

For ports to be classified as agile they need to be characterized by a continuous research to anticipate market trends and powerful information and decision technology and systems. Thus, each port would act as an information distribution centre within a communication network and each stakeholder should be able to make decisions *just-in-time* without wasting the time and money of the other stakeholders.

Before a port becomes agile, it has to be a *lean port* that moves cargo quickly and smoothly, delivering a service in line with market demand while eliminating waste within its processes. A lean port is one that as a business unit makes the best use of its available tangible and intangible resources and, by eliminating all sorts of waste in the physical and documentary/information processes related to the cargo and the different modes that service the port, delivers "perfect" customer service.

The concept of *lean port network* is particularly relevant whenever there is a need to develop multimodal transport systems that can compete efficiently and effectively with unimodal transport such as road, thus enabling the utilization of underused capacity of water and rail-based carriers. A minimization of costs associated with delays and wastes, the use of the best (in terms of cost and efficiency) transport chain given the quantity of cargo and distance to be covered, and a capability to respond to volatile demands caused by market uncertainty are expected from the development of relationships of lean port networks and inland terminals. This will provide an efficient transport network, particularly required by contemporary logistics (value) chains. Lean and agile ports are thus supposed to become flexible, responsive, adaptable, and knowledge centers.

In addition, three broad interdependent elements shape the efficiency of seaport terminals: capacity, regulatory requirements, and organisational relationships and type of control over container movements.

Capacity refers to the physical ability of the port terminal to store containers and to transfer the containers between ship and shore. This represents a fundamental issue because of the basic difference in capacity between ships and land modes, which makes storage essential, and also the need to turn ships around with the minimum delay. A very wide range of factors influence physical capacities that lend themselves to operational modeling.

Ocean freight shipping has to meet stringent *regulatory requirements*, most notably customs documentation, but also new security stipulations such as the International Maritime Organization's International Ship and Port Security (ISPS) code and the United States' 24 hour rule (detailed manifest information for US-bound traffic, requested by US customs, at least 24 hours before loading at a foreign port), labour stipulations, working conditions, hours of operation. These will have to be considered in the modeling operations.

Concerns with respect to *organisational relationships and type of control over container movements* raise a series of issues. On whose account does the container move? What are the conditions of pricing (port to port or door to door)? Who decides on routing? What are the relations between the parties involved in physical handling? All these issues will have to be considered in developing a fully-fledged model.

IV.4. ITS and Intermodality

Over the last few years, there have been a number of significant technological developments in the area of intermodal transportation. Most of these developments involved advanced location (positioning) and communication technologies which fall under the realm of *Intelligent Transportation Systems* (ITS).

The term *ITS* is generally used to refer to the newest technology, infrastructure, and services, as well as the planning, operation, and control methods to be used for the transportation of persons and freight. It was the obvious answer to the need to significantly increase the capacity of transportation systems without heavy investment in new infrastructure. ITS makes transportation systems more efficient through an integrated use of the latest developments in various areas, infrastructure and vehicle technologies, electronics, telecommunications, computing hardware, positioning systems, as well as advanced data processing and sophisticated planning and operation methods. In the beginning, ITS research, development, and investment focused on urban automobile transportation and a totally public organizational structure and management. It has now evolved to include all types and levels of transportation, persons as well as freight, for which private industries offer a variety of extended, adapted and targeted services.

Tremendous challenges and opportunities exist for ITS research, development, and business, particularly so in the area of freight transportation that, until recently, appeared relatively less prominently on the agenda of ITS stakeholders. Indeed, the development of Intermodal Freight ITS and the evolution of the intermodal freight transportation industry are closely related, particularly relative to the use of information and decision technologies in response to the tremendous shift in commercial and industrial practices of the last decade (Crainic and Gendreau 2003, Crainic 2004). We are interested in this section in surveying the most important studies and contributions of ITS technologies with respect to intermodal freight transportation.

Eckhardt *et al.* (2003) studied the integration of tracking and tracing systems in multimodal transport chains. The study was based on the Finnish freight transport telematics architecture, which covers different modes of transport including road, rail, water and air. The main objective of the study was to examine the needs and possibilities of integrating tracking and tracing systems of different transportation modes and companies.

The information flows considered comprises management data, track and trace data, and other data related to the goods orders, transport agreements, transport orders, dispatch notes and load specifications. For data transfer and transmissions, data networks and mobile telephone networks were used, as well as satellite communication (for some pilot projects), and Internet, due to its wide accessibility for both customers and firms. The most important and used tracking and identification techniques were barcodes and RFID (Radio Frequency Identification), the latter being considered more suitable for automated identification than barcodes because of its longer identification distance, no need for visual contact, and no need for the target to be stopped for the recognition.

As a result of this explorative research, it was found that the information considered as the most important by the users of tracking systems concerns *schedules*, in other words, the time when the goods are delivered. The real time location of the transport was found less significant. In fact, the tracking and tracing system, in the eyes of their potential users, should cover the whole supply chain and all or nearly all commodity groups and supply chain actors. In addition to this information, the tracking system should also contain property data of goods information for the control of the supply chain operation. Only ID information would move with the handling unit, and other information would be in the background system.

The biggest obstacles in the development of the tracking system are the lack of operating models and shortcomings of the standards. Rapid technical developments also add uncertainty to the introduction of systems. In spite of these obstacles, the development of systems should be started (Eckhardt *et al.* 2003). Indeed, on the basis of the logistics experts interviewed, Eckhardt *et al.* (2003) concluded that there is a need for an integrated tracking & tracing system. Some of the possible advantages could be the quick reaction to changes and minimization of the drawbacks caused by changes. Also it might be possible to decrease logistics costs. Standardization, common reference and easy connectability to other systems are the probable preconditions for such an integrated system.

Even if the tracing is successful with the help of some "in house" system, it is still difficult to interpret the information and decide whether the schedule commitments will be fulfilled or not. In fact, the information needed is about what *will really happen*, as compared to what was originally planned. An example is given by Dürr and Giannopoulos (2003): in case of a delay the user has to judge whether the cargo can catch-up later to the originally planned arrival time or what the new estimated arrival time (EAT) is; this illustrates the fact that interpretation and use of this kind of data is difficult indeed.

The data collection stage is the one in which the various data elements are collected with respect to the specific transport. In other words, it provides *the content* of the data transfer and transmission that accompanies a certain transport. The ways and means of collecting these data are many and vary from the various Tracking and Tracing equipment (Global Positioning System (GPS), or Global System for Mobile communications (GSM), or satellite based, etc.), to electronic tags, or more "manual" types of data entry into the computer, etc.

Similarly to freight intermodal transportation, the reduction of delivery lead time of the supply chain system constitutes one of the main priorities for the *air cargo logistics industry*. A massive amount of data is processed and communicated in-between parties, which may include packing and unpacking, transportation and handling, customs clearing, carrier bookings, warehousing, and other related activities such as insurance. At present, most service providers do allow their customers to access their *Internet cargo tracking systems* to search for correct consignments and related flight information. However, the lack of coordination between agents is making it very difficult for a customer to track down the complete status of a shipment (Li and Shue 2003).

It is traditional to examine Freight ITS according to the scope of the systems classified in two broad classes: *Commercial Vehicle Operations* (CVO) for system-wide, regional, national, or continental applications and *Advanced Fleet Management Systems* (AFMS) dedicated to the operations of a particular (group of) firm(s) (Crainic 2004, Crainic and Gendreau 2003). A major class of CVO projects, particularly widespread in North America, concerns *border crossing operations*. This area has acquired a sense of urgency and high priority following the terrorist attacks on the United States and the continuing terrorist threat. Ports have thus become prime targets for ITS and e-business projects with security issues as the driving objective. While the urgency has been primarily felt in the United States (Transportation Research Board 2002), border CVO systems are being developed worldwide. The main goal was and continues to be to clear drivers, vehicles, and cargo in order to speed up the passage of vehicles (trucks, containers, railcars) carrying manufactured and agricultural goods through the border inspection facilities, within the parameters set by the border control requirements in terms of security, immigration, illicit cargo, agricultural controls, etc.

Several security policies significantly increase delays at ports and border crossings and thus influence the efficiency of commerce and supply chains. Among others, the U.S. Customs *Container Security Initiative* (CSI) requires the inspection and pre-clearance of a certain proportion of containers *before* they leave the port of origin or the last major transshipment port. The U.S. Customs and Border Protection Agency also requires advanced transmission of cargo information for shipments destined for the United States. Systems are being deployed to mitigate the associated significantly longer delays. For instance, U.S. and Canadian customs commercial programs are being aligned (the *Free* *and Secure Trade* – FAST – program) to support moving pre-approved goods quickly across the border. The program is based on registering and pre-approving import/export firms (shippers), carriers, and drivers.

For ports and border ITS/CVO, as for most other ITS areas, the development of the "intelligence" part must accompany that of the hardware and the availability of information. Very few efforts have been undertaken in this area, however. The determination of the optimal number of containers to be inspected to satisfy the security requirements and to limit the delays in ports is an example of such a topic (Lee et al., 2003; Lewis et al., 2002, 2003). Many challenges and opportunities are also offered by the intense automation of container terminals in ports (Crainic and Kim 2005).

Several European research projects on transportation targeted these issues: EUROBORDER (1998), TR 1033 INTERPORT (2000), PROSIT (Schinas *et al.* 2002). The objectives of EUROBORDER consortium, whose work and results are presented over a time lag of 2 years and a half, were to identify bottlenecks of ports as nodal points in the intermodal transportation chain, as well as to identify informational, organizational and administrative requirements, concepts and functionalities to integrate ports into the intermodal transport chain including the Trans-European Networks. The methodology they have chosen was to perform case studies of a small number of ports, Piraeus, Bilbao, Oslo, and Helsinki. To analyse the effects of the EUROBORDER scenarios on transport chains, four network models, based upon a few real-life transport chains involving the Norwegian, the Spanish and the Greek scenario ports, have been deployed. By analysing these real cases, it was found that the port costs constitute a rather small part of the total generalised costs for a transport from door to door. The quantitative analysis performed, based on generalised costs, showed that even if prices and time are important, these two factors alone do not explain the choice of a transport solution. The quality of service provided (reliability, flexibility, security, etc.) was found to be equally or even more important. EUROBORDER (1998) concluded that port terminals must position themselves as an integral part of intermodal transport chains (hinterland-port-sea-port-hinterland), and that there is a big uncertainty characterizing information handling (internally and especially externally), due to the diversity of customers and their needs, lack of standards and lack of trained staff. Thus, three key issues were identified as the most important for ports to consider for improving their performances: *interaction* (to actively define a role and a business strategy), *integration* (to coordinate investments and operations to exploit synergies), and *information* (to manage the intermodal transport tasks and contribution to the overall goal).

The INTERPORT project was focused on exploring the use of automatic identification equipment for the monitoring of load units, vehicles and staff in port terminals. The issues addressed included costs, benefits, compatibility and practical aspects, as well as standardization. As the work performed was carried out in close co-operation with terminal operators of six ports throughout Europe (Piraeus Container, Piraeus Ferry, Thessaloniki, Oslo Color Line, Helsinki North Harbour, Bilbao MacAndrews). This gave to the application a wide operation perspective, with variations between terminals due to differences in traffic, service level requirements, organisation, decision making process, etc. A system architecture and a data model for intermodal transport were defined and results and experience gathered were summarized in a handbook (*INTERPORT user handbook*), which gives guidance on how to develop a system architecture, how to address standardisation and evaluation issues, and discusses problems related to implementation and organisation. Much of the standardisation discussion about automatic identification equipment has focused on the problem of tagging all containers in the world. In reality – especially in European short sea shipping – other transport units are more in use and far more dedicated to certain transport relations than overseas containers. This conclusion opened opportunities for starting the installations for specific flows, routes or customers and gradually expanding the system.

The PROSIT project constituted an effort to introduce telematics technologies in the traditional field of the shipbroker, with the general aim to improve the overall efficiency of waterborne transport by the use of advanced telematics solutions and to support an intermodal brokerage between the demand (industry and trade) and supply (transport industry) sides of freight transportation. The research was focused on exploring the possibilities of substitution of an actual commercial procedure with web-based tools. Four major scenarios, involving different states of technology, market needs, and organizational structures, served for the development of the project. The "applied" goal of PROSIT was to act as a test platform for research work in transport industry, environment, and European transport policy. After validation and evaluation of the overall objectives, PROSIT was found by its promoters to successfully support planning and control along the transport chain and improve co-operation and communication between the partners. It was not clear though wheatear it supported successfully an intermodal brokerage between demand and supply. The doubts lied mainly on the issue of business culture and the inherent resistance to changes.

In the same context of (electronic) intermediaries or "infomediaries", several firms address spot market transactions and long-term contract negotiations for truckload, airfreight, ocean and intermodal shipments. An interesting research work, conducted by Boyle (2000), examined indirect channels for freight transportation and the specific functions performed by existing intermediaries (brokers and forwarders). Frameworks have been provided for channel structure and strategic groupings. It was concluded that shippers would receive the greatest benefit from Business to Business (B2B) marketplaces and that Internet-based transportation systems present the best opportunity for value creation.

IV.5. Container Release-Time Forecasting

We are concerned here with the methodology for the development of a container release forecasting and dynamic update methodology. Various approaches could be considered, including econometric and time-series models, as well as dynamic learning methodologies, such as Artificial Neural Networks (ANN). Econometric and time series models are more data demanding and are difficult in a dynamic context were data is continuously updated. They are estimated using a particular sample and even if they perform very well for this sample, they do not adapt easily to changing structural conditions. Furthermore, their adaptation/updating procedure may be difficult to implement for non-specialist end-users. Finally, such methodologies have not been developed for generating forecasts to be used in real-time planning models.

Dynamic learning methodologies seem to correspond better to our purposes. Considering an *Artificial Neural Network* (*ANN*) approach (Haykin 1994, Potvin and Smith 2001), the forecast updating process is robust (graceful degradation property) and can be performed in a flexible way (e.g., on a daily basis). Another advantage of this approach is its adaptability to change. The neural network can be modeled such that it takes into account the dynamics and evolution of a time-dependent system like ours. It can also be less data demanding compared to the econometric or time-series approaches.

A comprehensive review of the literature about neural networks in transportation was published in 1995 (Dougherty 1995). Over 40 papers published between 1990 and 1995 were reviewed, illustrating the explosion of interest for transport studies that used neural networks during that period. The distribution of the reviewed papers, with respect to the application subject area and to the ANN paradigm used is given, respectively, in Tables 1 and 2. In 1997, Dougherty also initiated the publication of a special issue of *Transportation Research Part C* on applications of neural networks in transportation. In its editorial, Dougherty (1997) indicated that "the pace of development in the theoretical aspects of neural computing has begun to slow down". This was seen as good news for the possible implementations of artificial neural networks, as ANN were entering a period of consolidation, characterized by broader acceptance from the scientific community, as a relatively stable, proven technology, rather than a fast-changing one.

Subject area	Number of papers
Driver behaviour/autonomous vehicles	12
Parameter estimation	7
Pavement maintenance	6
Vehicle detection/classification	5
Traffic pattern analysis	5
Freight operations	4
Traffic forecasting	4
Transport policy and economics	2
Air transport	2
Maritime transport	2
Submarine vehicles	1
Metro operations	1
Traffic control	1
Total	52

Table 1. ANN Subject Area Distribution (Dougherty 1995)

Paradigm	Number
Backpropagation	36
Learning vector quantisation	7
Adaptive resonance theory	4
Self-organising map	2
ADALINE	1
Hopfield	1
Basis functions	1
Counterpropagation	1
Boltzmann machine	1

Table 2. ANN Paradigm Used (Dougherty 1995)

Transportation research is also currently entering a period of rapid change, a period which presents a unique opportunity for new styles of data analysis in order to meet the new needs for efficiently and comprehensively exploring large databases for patterns and relationships against a background of data uncertainty and noise, especially when the underlying database is of the order of gigabytes. *Computational intelligence technologies* such as the computational neural networks show the potential for a new paradigm in transportation research, providing researchers with rich and interesting classes of novel data driven methods and techniques applicable to a wide range of domains (Fischer 1998, Sadek *et al.* 2003).

In this context, neural networks are seen as a supplementary method for data analysis rather than a substitute. It is a useful method for relatively assumption-free reconnaissance of large data sets (Shmueli *et al.* 1998). They operate directly on the data, avoiding the necessity of a priori assumptions of a structural model with all consequential constraints.

Quality and quantity of data are widely recognized as important issues in the development of neural network models (Lyons *et al.* 1998). Performance is sensitive to the level of detail with which a real world scenario is described to the neural network model. Insufficient detail can make examples from different classes indistinguishable. Excessive detail increases the extent of the input environment space, which could result in over-characterization of a part of the input environment which is more difficult to interpret during training.

In a recent paper by Lam *et al.* (2004), neural network models are proposed and developed for forecasting 37 types of freight movements in the Hong Kong port (cargo throughput from 2002 to 2011). The data used for model development consists of the history records (1983-2000) of freight movements and explanatory factors. Based on the results obtained, forecasts made by the proposed neural network models are said to be more conservative, more reliable, and more comparable to reality, compared to the use of an econometric analysis approach.

Al-Deek (2001b), Klodzinski and Al-Deek (2003), and Sarvareddy *et al.* (2003) present neural network approaches for a problem very close to ours: predicting the levels of cargo truck traffic moving inbound and outbound at seaports, for developing freight planning models with specific applications to seaports. Regression analysis and back-propagation neural networks (BPN) were compared with respect to their respective performances in short-term forecasts of truck trips at several ports in Florida (Port Everglades, Port of Tampa, Port of Miami, Port of Jacksonville, and Port Canaveral).

Al-Deek compared the use of neural networks to multiple regression analysis for application to freight modeling (Al-Deek 2001a, 2002) and developed a methodology for applying Artificial Neural Network modeling which makes use of vessel freight data for computing daily truck volumes generated by seaport vessel freight activity. The neural network model results were found significantly more accurate than regression analysis for the Florida ports for which were they were applied.

The primary factors affecting truck and rail volume were found to be the amount and direction of cargo vessel freight, commodity type (bulk, break bulk, or liquid bulk), and weekday of operation. Five ANN models were developed successfully for the selected Florida ports. The BPN algorithm was used for all of them, but it required an elaborate process and was unique to each port. Other modeling techniques were investigated and a Fully-Recurrent Neural Network (FRNN), trained with the Real Time Recurrent Learning (RTRL) algorithm, was selected to develop the model for one of the ports, in order to make a comparison with the BPN model. FRNN is dynamic in nature and was found to have the ability of relating the storage time of the commodities in the port with truck trip generation. It could adequately model the current conditions but failed to forecast truck volumes in the future. The FRNN was found to require more data for forecasting than BPN, but it is supposed to produce more accurate results when the necessary quantity of data is available for training.

Malavasi and Ricci (2001) address several aspects of the process of simulating stochastic elements in railway systems using self-learning processes. Their study tested the use of systems capable to reproduce events without requiring a previous definition of the operating rules. For the development of the neural network model, they identified six life cycle phases, as follows:

1. Definition of goals and requirements of the system;
2. Check and pre-processing of the data;
3. Selection of the neural processing tool best suitable for the definition of the system as in the first cycle (to be iterated, if necessary, after collecting a wider set of data, after changing the tool or after partitioning the problem);
4. Training of the model and testing it on out-of-sample data;
5. Application of the network;
6. Maintenance by means of progressively updated data sources.

Very encouraging results with respect to the performances of forecasting models based on neural networks compared to other (conventional) methods of prediction are presented in papers by Dharia and Adeli (2003), Huisken and van Berkum (2002), Sayed and Razavi (2000), who are using a neurofuzzy approach, Abdelwahab and Sayed (1999), Faghri *et al.* (1999), who tested and compared five different architectures of backpropagation against three different regression models and found that the worst ANN prediction was better than the best regression one, Park and Rilett (1998), who found that the modular ANN gave the best overall results compared to existing forecasting techniques (Kalman filtering, exponential smoothing, historical profile, real-time profile), and Bansal *et al.* (1998), to name just a few.

The specific problem that we are interested in for our present project was not directly treated in the literature. Nevertheless, similar problems were approached using a neural network model (Al-Deek 2001b, Klodzinski and Al-Deek 2003, Sarvareddy *et al.* 2003), which allowed us to contemplate the development of a container release forecasting and dynamic update methodology as feasible, provided the necessary data was to be made available.

V. Data Issues

In order to tackle the problem as described in the preceding sections, the first concern was the information available to CN regarding the release of containers from port terminals. This was aimed to set up the problem characteristics and identify the data currently available or that could be made available in the future (on the basis of recommendations). As indicated earlier, the project does not aim to optimize the chain of operations within the port. However, we undertook the initiative to build an image of this chain, as well as of the associated information and decision chains, in order to enhance the understanding of the process and to permit an assessment of potential benefits that could be derived by applying ITS techniques to it. This initiative was conducted through working meetings and particular interviews with the CN team and port managers and specialists.

The study focussed on the so-called Eastern intermodal service network with main terminals in Chicago, Toronto, Montreal, and Halifax, and ports in Montreal and Halifax (see Figure 6).

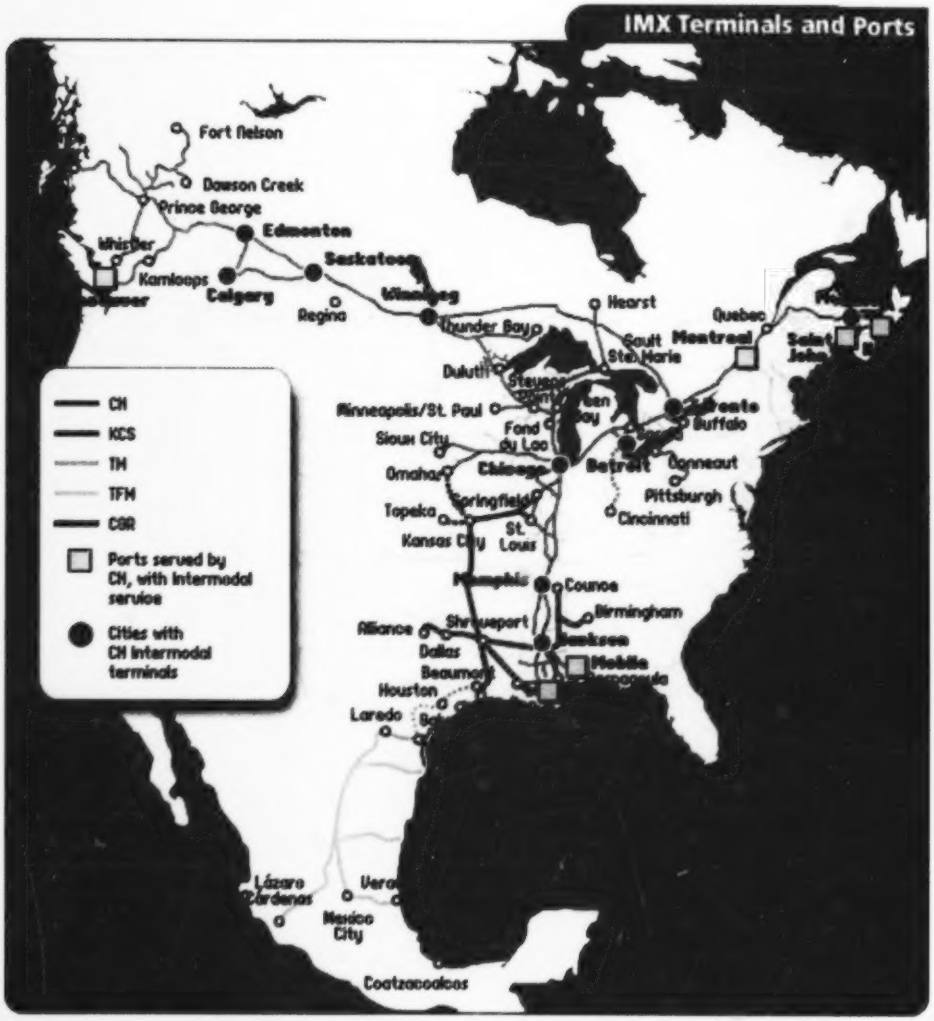

IMX Terminals and Ports
Fort Nelson
Dawson Creek
Prince George
Edmonton
Whistler
Kamloops
Saskatoon
Calgary
Regina
Winnipeg
Thunder Bay
Hearst
Quebec
Montreal
Sault Ste. Marie
Duluth
Stevens Point
Green Bay
Minneapolis/St. Paul
Fond du Lac
Sioux City
Toronto
Buffalo
Omaha
Chicago
Detroit
Conneaut
Pittsburgh
Cincinnati
Topeka
Springfield
St. Louis
Kansas City
Memphis
Counce
Birmingham
Alliance
Shreveport
Jackson
Dallas
Beaumont
Mobile
Houston
Laredo
Lázaro Cárdenas
Mexico City
Coatzacoalcos
Saint John
CN
KCS
TM
TFM
CSR
Ports served by CN, with intermodal service
Cities with CN intermodal terminals

Figure 6. The CN Intermodal Service Network
(www.cn.ca/specialized/ports_docks/en_KFPorts.shtml)

Time considerations

The simulation, prediction, and service-adjustment models are time dependent, that is, the models explicitly include the time dimension. This allows representing adequately the operations of the system and the time dependencies among the associated activities and events. Such an approach requires a definition of "time", in terms of both planning horizon and time unit.

The CN schedules repeat every day. The "day" therefore appears as the natural time unit for the prediction and service-adjustment models. The simulation model is built according to an event-driven strategy, i.e., the simulation advances to the time of the next scheduled event. Consequently, actual schedules are simulated and the time unit is the minute.

The planning horizon, i.e., the length of time considered, is specific to each model. It varies from a few days for the prediction model, to a few weeks for the service-adjustment model, to several months for the simulation. See Sections VI, VII, and VIII for additional details.

Data collection

The process of collecting and organizing (pre-processing) the data intended to feed the prediction and service-adjustment models, as well as the simulator, was a critical phase in our study development. The fact that real data is used for training and validating the NN is a very challenging issue, from multiple perspectives.

The process of forecasting demand has to take into account all the information that could be obtained, which implies also the latest information available in the system, in a dynamic manner. Thus, information about ship (containers) arrival in port/terminal is important, since there is a strong correlation between this data and the corresponding container release time (delay). Ship/containers arrival data concerning the quantities, types of containers, their desired destinations, etc., and ship arrival dates were thus needed. Primarily, this type of information was expected to become available from the ports themselves (Montreal or Halifax). Actually, some CN data concerning ship arrival dates and total quantities of containers (in FEUs) contained on them were made available by CN starting July 2005.

The type of data collected depended on the type of data identified and available at CN. Desired data availability issues had to be explored and data requirements had to be specified. Two types of data were primarily required for the project: CN data, which were made available for use in the project, and port data. The latter was not available for this study, but visits and interviews in the ports of Montreal and Halifax contributed to our understanding of the real problems port terminals are confronted with.

The data initially requested from CN concerned the chronological information about CN containers, individually, from the departure from the origin port, to the arrival of the ship in port, and the container's day of transportation by CN from port to destination (should not all this information be available, partial information was also specified to be of interest for our study). For each container (C), were needed:

- Characteristics: ID – identifier (if any), S – size, T - type, O – origin, D – destination, P – priority (if any), etc.;

- Scheduled arrival date in port (SAD);
- Actual arrival date in port (AAD);
- Exit date from port, before being put on a CN train (EDP);
- Departure date on CN train toward destination (DDCN)
  - Train and block information;
  - Arrival date at destination and intermediary stops.

This information was specified as needed for a long time interval (2-3 years back), in order to allow reliable forecasts. Samples (on shorter periods) were specified to be of use for the design and validation of the forecasting model. As samples were finally to be used, they were needed to cover characteristic periods during the year: seasons, holidays, ups, downs, etc.

CN provided several sets of real data characterising the studied intermodal services over the period of time from January 2005 to July 2005 (two months in winter time and two months in summer time), for the cities of Chicago, Toronto, Montreal and Halifax. The data is constituted by record of containers that were transported on intermodal CN services between those cities over the specified periods of time. These services were operated daily over the network illustrated in Figure 6.

The description of a container takes into consideration some of its physical characteristics, such as its length, weight and type. It also takes into consideration geographic characteristics such as its origin, destination, the terminal where the container is put on a car ("ramped"), and the dates associated to the container corresponding operations (container arrival at origin, container arrival to destination, container ramping date). It takes also into consideration the equipment that carries it, and thus, in the description, there are fields corresponding to information related to cars and trains. Rail cars are classified with respect to their physical characteristics (dimensions, maximal capacity, etc). They may be formed of one or several platforms linked together.

As requested, the data files contained lists of all individual containers (in the data table, one record per container) that were transported by a CN intermodal service during the specified period of time. The set of attributes and their possible values describing the containers are listed below.

A selection of these attributes was required for the use with the neuronal network-based demand model (container release time forecasting) and another selection of these attributes was necessary for designing and feeding the simulator, as it will be further detailed in this section. The set of attributes characterising one container – one entry of the data file – is detailed in the following list:

- *EquipmentInitial* This attribute is used (in conjunction with the EquipmentNumber) to uniquely identify the container. It consists of a four-position alphabetic reporting marks;
- *EquipmentNumber* This attribute is used (in conjunction with the EquipmentInitial) to uniquely identify the container. It consists of a six-digit number;

- *LoadEmpty* This attribute is a tag indicating whether the container is loaded (L) or empty (E);
- *CarKind* This attribute is a code assigned by CN to group containers with similar characteristics;
- *Origin* This attribute is the abbreviation corresponding to the geographic origin of the container. The abbreviation is divided into three parts, each one containing three characters. For example, MONTASYAR is interpreted as MON (for Montreal), TAS (for Taschereau) and YAR (for yard);
- *Destination* This attribute is the abbreviation corresponding to the geographic destination of the container and has to be read in the same way as the origin;
- *OperatingCity* This attribute has to be interpreted as the container geographic destination when the Destination field is missing;
- *PlanNumber* This attribute is a code indicating the intermodal service supplied. For example, "27" stands for "ramp to door - railway owned equipment", "42" stands for "door to ramp - steamship line equipment";
- *ASStation* This attribute is the assignment station, which means the CN station where the container was assigned to a rail car to be ramped on;
- *AS-OR* This attribute is the date corresponding to the assignment operation;
- *RMStation* This attribute is the CN station where the container is ramped, which means where the container is put on a car;
- *RM-OR* This attribute is the date corresponding to the ramping operation;
- *IGStation* This attribute is the "In-Gate" station, which means the station were the container was received by CN from another mode of transportation; this applies only on non-ramped containers, and thus not for traffic coming from ports (containers from ports arrive always already ramped to CN);
- *IG* This attribute is the date corresponding to the "In-Gate" operation;
- *TDStation* This attribute is the CN station of train departure;
- *TD-OR* This attribute is the date corresponding to train departure;
- *TAStation* This attribute is the CN train arrival station;
- *TA-DS* This attribute is the date of train arrival in the CN destination station;
- *AtchEquipmentInitial* This attribute, used with the AtchEquipmentNo, is meant to uniquely identify the rail-car on which the container is ramped on. It consists of two to four alphabetic marks;
- *AtchEquipmentNo* This attribute is used with the AtchEquipmentInitial to uniquely identify the rail-car on which the container is ramped on. It consists of one to six digit number;
- *Shipper* This attribute is an abbreviation describing the container shipper and consists of a maximum of twelve characters;

- *OSLength* This attribute is the outside length of the container in feet;
- *TrackNo* This attribute is a number describing a zone and a track at a station that is going to be used by the train that will carry the container;
- *EquipmentWeight* This attribute is the container weight, excluding the weight of the empty container (zero is a possible value);
- *EquipmentWeightUnit* This attribute is a tag describing the weight unit: 'L' stands for pound, 'K' stands for kilograms;
- *TrainID* This attribute identifies the train that will carry the container. It is divided into four parts: the train service type (Q stands for intermodal), the train number (consisting of three digits), the region (consisting of one digit), and the section (consisting of one digit). The last two digits are the date the train was scheduled to depart from the origin station;
- *PlatformCode* In the CN data, we refer to a unitary rail-car as a "platform". This attribute is a letter between A and E describing the position of the platform on the rail-car that will carry the container;
- *PlatformPosition* This attribute is the position of the container on the platform: "1" and "2" stand for the two positions at the bottom of the platform, and 'T' stands for the position of the container on top (20 feet containers must travel "at the bottom" of the car; when two twenty feet containers are at the bottom of a car, "1" and "2" indicating which one of the two possible positions is used; when a 40 feet container is put on the bottom of a platform, the position identifier "1" is used only).
- *CAR_KIND* This attribute describes the group of rail-cars to which belongs the car that carries the container. Here is the description of the "car" types. LCap and HCap stand respectively for "low capacity" and "high capacity" cars. The rail-car types used for our simulation purposes are given in the list below:

LCap - **PC3** CONVENTIONAL CAR COFC ONLY - SINGLE STACK ONLY - MAX 3 UNITS
LCap - **PC4** CONVENTIONAL CAR COFC ONLY - SINGLE STACK ONLY - MAX 4 UNITS
LCap - **PD3** CONVENTIONAL CAR TOFC/COFC - SINGLE STACK ONLY - MAX 3 UNITS
LCap - **PD4** CONVENTIONAL CAR TOFC/COFC - SINGLE STACK ONLY - MAX 4 UNITS
LCap - **PRR** STEEL WHEEL RAILSETS FOR CARLESS TECHNOLOGY
LCap - **PS2** WELL CAR COFC ONLY - DOUBLE STACK CAPABLE - MAX 2 UNITS
LCap - **PT1** CONVENTIONAL CAR TOFC ONLY - 1 HITCH
LCap - **PT3** CONVENTIONAL CAR TOFC ONLY - 3 HITCHES
LCap - **QC2** MULTI-PLATFORM COFC ONLY - SINGLE STACK ONLY - 2 WELLS/PLATFORMS
LCap - **QC5** MULTI-PLATFORM COFC ONLY - SINGLE STACK ONLY - 5 WELLS/PLATFORMS
LCap - **QD2** MULTI-PLATFORM COFC/TOFC - SINGLE STACK ONLY - 2 WELLS/PLATFORMS
LCap - **QD3** MULTI PLATFORM COFC/TOFC - SINGLE STACK ONLY - 3 WELLS/PLATFORMS
LCap - **QD5** MULTI PLATFORM COFC/TOFC - SINGLE STACK ONLY - 5 WELLS/PLATFORMS
LCap - **QS5** MULTI-PLATFORM COFC ONLY - DOUBLE STACK CAPABLE - 5 WELLS
LCap - **QT2** MULTI-PLATFORM TOFC ONLY - 2 PLATFORMS
LCap - **QT3** MULTI-PLATFORM TOFC ONLY - 3 PLATFORMS
LCap - **QT5** MULTI-PLATFORM TOFC ONLY - 5 PLATFORMS
1/3 – HCap and 2/3 – LCap of the total amount of cars in the system:
QS3 MULTI-PLATFORM COFC ONLY - DOUBLE STACK CAPABLE - 3 WELLS
HCap - **PB3** CONVENTIONAL CAR TOFC/COFC - DOUBLE STACK CAPABLE - MAX 3 UNITS
HCap - **PS3** WELL CAR COFC ONLY - DOUBLE STACK CAPABLE - MAX 3 UNITS
HCap - **PS4** WELL CAR COFC ONLY - DOUBLE STACK CAPABLE - MAX 4 UNITS

HCap - **QB3** MULTI-PLATFORM COFC/TOFC - DOUBLE STACK CAPABLE - 3 WELLS/PLATFORMS
HCap - **QB4** MULTI-PLATFORM COFC/TOFC - DOUBLE STACK CAPABLE - 4 WELLS/PLATFORMS
HCap - **QB5** MULTI-PLATFORM COFC/TOFC - DOUBLE STACK CAPABLE - 5 WELLS/PLATFORMS
HCap - **QS4** MULTI PLATFORM COFC ONLY - DOUBLE STACK CAPABLE - 4 WELLS

- *FEET* This attribute is the outside length of the rail-car to which belongs the platform carrying the container;
- *TYPE* This attribute is the type of rail-car to which belongs the platform carrying the container;
- *Platforms* This attribute is the number of platforms characterising the corresponding rail-car;
- *Stack* This attribute is the car type: "DOUBLE" stands for "double stack", "CONV" for "single stack".

Data understanding, analysing and pre-processing

In order to feed the simulator, we needed data for all the demand points in the network. This data was requested to span a representative time window – 2 months in winter time and 2 months in summer time. Data had to contain, for each container:

- In-gate dates (to be as close as possible to the real demand);
- Train departure date (time), rail-car characteristics;
- Container characteristics: origin-destination, weight, loaded/empty.

In order to design and implement the container-release forecasting model, after several exchanges with CN personnel, the initial queries for the data files were modified such that the current attributes characterising the available data sets are the following:

- Equipment type (loaded or empty),
- Rail car type used,
- Origin of the container,
- Destination of the container,
- Ramping station and date,
- Train departure station and date,
- Owner of the container,
- Shipper of the container,
- Outside length of the container (containers were classified in 20 feet and 40 feet or more containers).

From the data files provided by CN, we have extracted and formatted the information necessary to build the necessary input files both for the neuronal network and the simulator. The needs being different for the two approaches, we will describe in two different sub-sections the expected information for each approach, and the work that has been done to extract this information. Note that for the simulator, extracting the real demand from the CN demand information was not only necessary to run simulations with real demand over the considered period of time, but was also the first step to deduce statistical distributions to be able to generate demand from random distribution over any period of time that we wish to simulate.

Demand generation for the container-release forecasting model

For the Neuronal Network based forecasting model, we need information about containers movements in the ports of Montreal and Halifax. More precisely, for each port, we need, for each day, the number of twenty feet containers and the number of forty feet containers ramped in the port, the number of low capacity cars and the number of high capacity cars leaving or entering the port.

To obtain this information, we need to process the original container information as follows. For each container, we stock three characteristics: the container type (twenty feet long or forty feet long), the type of rail-car carrying the container (the field CAR_KIND) and the tag of this rail-car (the concatenation of the fields AtchEquipmentInitial and AtchEquipmentNo). We classify all the containers in two ways, one by origin (in fact, by ramping since the Origin field is missing in many cases), destination and ramping date, the other one by origin (ramping station), destination and arrival date. Those two classifications allow us to compute the required values for the neuronal network based forecasting model.

We process the data for each port as follows. We consider each day of the period of time covered by the data. For each day, we count the number of twenty and forty feet containers that have the port as origin in the first classification (the one made by ramping date). To compute the number of cars leaving the port, we also use the first classification.

We proceed in the following way for each type of container, twenty feet long and forty feet long: for each day, for each container, we check the tag of the group of cars that carries the container. If we have not already considered this tag for the current day, we count the number of cars corresponding to the type of the group of cars carrying the container. If we have already considered this tag, we ignore that container. In this way, we make sure to count only once each car leaving the port. To compute the number of cars entering the port, we proceed in the same way, using the second classification (the one made by arrival date), since we are interested in the cars that arrive in the port each day.

Demand generation for the simulator

For the simulator, the expected information obtained from the original data is much more disaggregated. More precisely, we want, for each day of the period that is considered, an enumeration of the containers that were ramped that day. For each container, we are interested in two characteristics, its type (twenty feet long or forty feet long) and its weight. The weight characteristic will allow us to classify the container during the simulation according to four possible options: empty, light, medium or heavy. Threshold variables in the simulator define those four possible types for the container weight.

To obtain the desired information from the original, ***observed***, data, we classified all the containers according to their ramping date, their origin (ramping station) and their destination. For each container, we stocked its type and weight. Then, from this classification, we enumerated, for each day and for each Origin-Destination combination of the block plan, the containers that were ramped that day and that had this combination as origin and destination.

The next step involved the derivation of a statistical characterization of demand for each O-D pair in order to be able to randomly generate containers to be transported during the course of the simulation. More precisely, we wanted to find suitable distributions that we could use to generate demand, as needed in scenarios for the simulator. These distributions would have to reproduce the main characteristics of demand while remaining simple to understand and handle.

A first step in demand modeling involved the way differences in container type and weight would be handled. To avoid being faced with having to fit distributions for flows with small positive demands, we decided that flows of all types and weights for each O-D pair should be aggregated to produce an O-D demand, the idea being that the specific characteristics of containers could be generated in a second stage, on an individual basis. From that point on, we therefore focused on fitting distributions for the daily demand of all containers for each O-D pair.

The next important question we had to deal with was whether or not demand displayed clear, identifiable cyclic patterns, for instance with respect to the day of the week. We examined the patterns through time of all O-D flows and came to the conclusion that no noticeable "day-of-the-week effect" exists: basically, there seems to be comparable demand on each day of the week (though with sharp day-to-day variations). The observation of flows also revealed a surprisingly high occurrence of days with zero demands for any specific O-D pair. We thought this observation was important in the context and should be reflected in any distribution that would be fitted to the observed data. Furthermore, histograms of observed demand values revealed that daily O-D demands did not follow any kind of bell-shaped distribution (e.g., Normal or Poisson). For simplicity sake, we decided that fitting two parameters of the observed distributions would be adequate: the probability of having zero demand p_0 and the average demand d_{avg}. For non-zero demands, we use a uniform distribution between 1 and a value d_{max}, which is chosen to fit the average (i.e., d_{max} is such that $(1 - p_0)(1 + d_{max})/2 = d_{avg}$).

To prepare data for the simulator, we thus proceed as follows. For each Origin-Destination combination, we compute the probability of having zero containers, which is the proportion of days without demand over the complete period. We also record the maximum number of containers observed (for comparison purposes with the computed d_{max} value) and we compute the proportion of each type of container, twenty feet long and forty feet long. For each type of container, we determine the proportion of empty containers, light containers, medium containers and heavy containers.

With this statistical information, we can easily generate demand information over different period of times, using, for each day of the period, the probability of having zero demand and a uniform distribution between one and the maximum number of containers observed for the concerned O-D combination. When containers are generated for a specific day and a specific O-D combination, the proportions of container types and container weight types are used to assign different characteristics to the containers.

Conclusions

To generate the desired information, both for the neuronal network and the simulator, a great deal of previous work needed to be done to clarify the original information before it could be used. For example, we first made a dictionary to associate every origin or destination in the original data to a corresponding terminal of the CN network. We also had to determine the validity of each container, a container being declared valid when a minimum number of characteristics met some criteria.

Analysis of the results produced by the simulator highlighted some of the pitfalls involved in the preparation of the data. A critical one was that data was missing for some periods that could not be ascertained a priori, but had to be determined for each O-D pair on an ad hoc basis: in the screening process we implemented, any period of 7 consecutive days without demand (for a specific O-D pair) was considered to indicate a block of missing data. Obviously, there might be occurrences where this absence of recorded traffic corresponds to a real lack of demand, which leads our procedure to slightly overestimate demand, as was confirmed by some discussions with CN staff. This situation could, possibly, be addressed by working with data corresponding to significantly longer periods. Note that such an approach could introduce different biases, however, by covering, for example, non-homogeneous periods.

Another issue, which cannot, unfortunately, be easily remedied, is the fact that observed data only tells us what demand was accommodated historically by CN, not what the real (latent) demand was. This type of situation is encountered in many transportation applications and could only be alleviated by the implementation of a much more sophisticated system to record and track down demand as it is expressed. Such a system would require recording systematically data that is not currently collected by CN (e.g., requests of customers trying to book slots for containers), but even if this were done, we might still have difficulties to properly gauge the actual demand process. Moreover, the cost of such a system could not be easily justified in relation to day-to-day operation of the system. A study limited in time and involving all appropriate CN personnel (operations, sells, IT, etc.) appears more practical, even though it is not clear how the necessary financial support could be mustered.

The previous discussion stresses the importance of obtaining reliable data. It also emphasizes the general fact that more thought and attention must be devoted to the estimation of demand for intermodal transportation.

VI. Container Release Time Model

The scope of this part of the project was to obtain reliable predictions concerning the release times (by type and quantity) of containers available for rail transportation, on a daily basis.

The operations of intermodal port terminals are characterised by important uncertainties. Ship arrivals in port may be off-schedule and pre-planned terminal operations have thus to be revised. Weather conditions may affect terminal operations, e.g., snow storms or very low temperatures may cause important delays in container handling. The availability in empty cars on any given day will also impact operations (e.g., if the number of High and Low Capacity cars needed for loading the given demand is not available, some of the containers that should depart cannot be "ramped" and sent to the corresponding rail transportation service, which implies also an overhead in demand for the next day). The day of the week impacts the labour force availability since, for example, this availability is diminished or is more costly for the Terminal Management during week ends.

Ports constitute main continental gateways, in general, and for North-America and Canada, in particular. Customs and immigration controls are therefore an integral and critical part of the process containers must go through. In the contemporary context of tight security, these activities may add significant complexity and delays to port operations, which makes container release times even more difficult to predict.

Most intermodal (port) terminals have the "same" general "modus operandi" but important differences may be observed in organisation, labour force policies and relations, expenses budgeting and management, and so on. Differences may also be engendered by the type of traffic port terminals are dealing with: origin of containers, type of containers (dimensions, refrigerated or not, etc.), frequency of ship arrivals etc. Any predictive methodology must therefore be adapted to each particular case. This is our goal in this part of the project.

As indicated in Section IV.5, dynamic learning methodologies, such as *Artificial Neural Networks* (*ANN*s) seem to be appropriate for our goals. The forecast updating process is robust (*graceful-degradation* property) and can be performed in a flexible (i.e., different input data) and repetitive (e.g., on a daily basis) way. Another advantage of this approach is its adaptability to change. The neural network can be modelled to take into account the dynamics and evolution of a time-dependent system (recurrent neural networks) and may be set up to automatically and dynamically update itself.

ANNs can be seen as excellent tools for non-linear modelling of complex systems by *learning*. They may be seen also as a non-linear extension of statistical methods. *Supervised learning* is performed based on a set of input-output examples that are presented to the neural network. The input data is fed to the neural network, which adapts its parameters (based on well defined algorithms) in order to minimize the error between its output and the expected output (the one corresponding to the input data). Thus, after a certain number of training epochs (cycles), the network should (when the performances are good) present the capacity of *generalization*, which is the ability to compute satisfying outputs from inputs that were never presented to the network before. The validation patterns (that do not belong to the training set) do not have to exceed the ranges of the inputs and outputs variables of the learning patterns.

Description of the neural network methodology approach

A set of meetings with CN representatives allowed us to understand the cornerstones of the development of a demand model for the particular case of the current study. Figure 7 synthesizes the activity chain and the corresponding data flow.

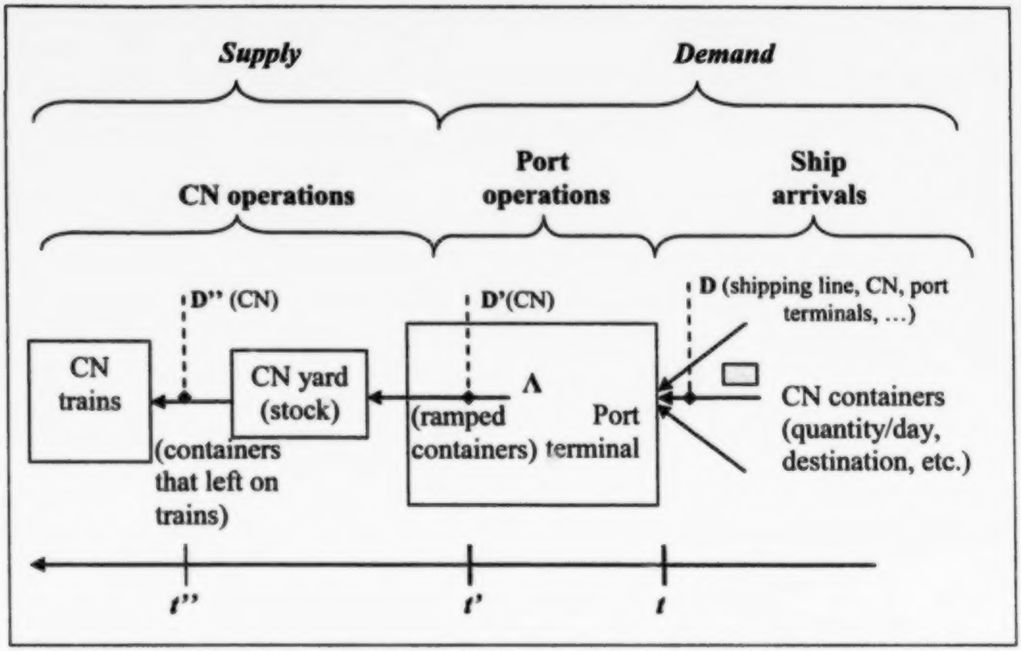

Supply
Demand
CN operations
Port operations
Ship arrivals
D'' (CN)
D'(CN)
D (shipping line, CN, port terminals, ...)
CN trains
CN yard (stock)
Λ
(ramped containers)
Port terminal
CN containers (quantity/day, destination, etc.)
(containers that left on trains)
t''
t'
t

Figure 7. Port Activity Chain and Corresponding Data Flow

The observed demand for CN services is represented by the quantities (and all associated characteristics contained in the individual waybills) of containers leaving the port (on rail cars) and entering the CN yard for selection and loading on trains. This data (**D'**) is available at port exit, at time (day) *t*. In the technical vocabulary of CN, these are the "ramped" containers.

Data denoted **D''** (ramped containers leaving the CN yard on train) is not adequate for forecasting/updating purposes because it is incomplete (censored), due to the selection process during which decisions are made about which containers are to be loaded and which containers are to be retained for a delayed (e.g., next day) departure. Port operations denoted (**Λ**) represent the set of all activities taking place in the port intermodal terminals, such as unloading containers from ships, in-terminal handling, custom clearance and CFIA controls, etc.

The purpose of the demand model is to "learn" the dependence that exists between the data related to the arrival of ships in ports (data denoted ***D*** in Figure 7) and the data related to the released containers once the port terminal operations have been performed (data denoted ***D'*** in Figure 7). This last is considered as the *real demand* data, because only this type of data, related to the already *ramped* containers at the exit of the port terminal, could be made available by CN. The containers not yet ramped, but available in the terminal, are not registered in the data transmitted to CN by the terminal. This discussion emphasises the crucial importance the collection and pre-processing of data has on the design and implementation of the demand model, as presented in Section V.

Data sets definition and model calibration

We have obtained and worked with data files containing historical information for 5 months (from January 1 to May 31, 2005) concerning the containers ramped in the Port of Montreal (a total of 80,418 records were analysed and treated).

Expected Arrival Times (EAT) of ships in port (data denoted **D** in Figure 7) were made available and could be successfully used, but detailed ship contents (total units per customer/per destination) were not accessible. Nevertheless, we could have information about global figures concerning the total amount of containers (in FEUs) arriving on each ship (aggregated information about the demand) for the same period (from January 1 to May 31, 2005 for the Port of Montreal).

Several sets of data were analysed and pre-processed, in order to be fed to the ANN, according to the following general requirements (notations refer to Figure 7). The input vectors consisted of:

- Historical daily data of type **D** : arrivals of ships in port and the freight quantities (FEU) corresponding to the pre-defined time horizon (sliding window of width ***n***): *day* ***j****, day* ***j*-1**,..., *day* (***j-n+1***);
- Other auxiliary information characterising the context (temporal, environmental, economic etc.) of the updating process, which would allow a better correlation between the output and the corresponding input;

The output vector (the data used as *updated forecasts*, once the ANN training process was completed and validated) consisted of historical data of type **D′**, corresponding to the current day or a given day in the future (***f***): *day* ***j****, day* ***j*** + **1**,..., *day* ***j*** + ***f***. The choice of the desired type of result (what quantities, what representation) was made according to the corresponding information needed as input for the simulation model.

Figure 8 displays the structure of a generic ANN. The definition of the *input layer variables* implies the choice of different data sets, for training, tests, and validation of the neural network model. The chosen information constitutes *representative* independent variables as follows:

- Month (from January to May);
- Day of the week (from Monday to Sunday);
- Terminal identifier (three different terminals for the Port of Montreal); as one of the three terminals was closed during the fourth month of our study-data, a decision was made not to use this information in the training process;
- Vessel identifier (15 different vessel identifiers for the Port of Montreal);
- Relative ship delay with respect to the Scheduled Arrival Day (in days of delay);
- 7 entries corresponding to a 7 days sliding window providing information on the quantities (in FEUs) of containerised merchandise arriving in the Port of Montreal daily (by one or several vessels arriving the same day); a choice was made to represent as an independent event (a single pattern) the individual ship arrival in port but to represent the total quantity of containers arriving as a cumulated daily information – this information is used for each of the seven days of the sliding window;

- 2 entries corresponding to the number of rail-cars having left the Port of Montreal the day before the current (j-1), by rail-car type (one entry for the number of High Capacity and another entry for the number of Low Capacity rail cars leaving the port the previous day); this was the only consistent data available in the actual system to give some information related to the available number of rail-cars in-port for a given day.

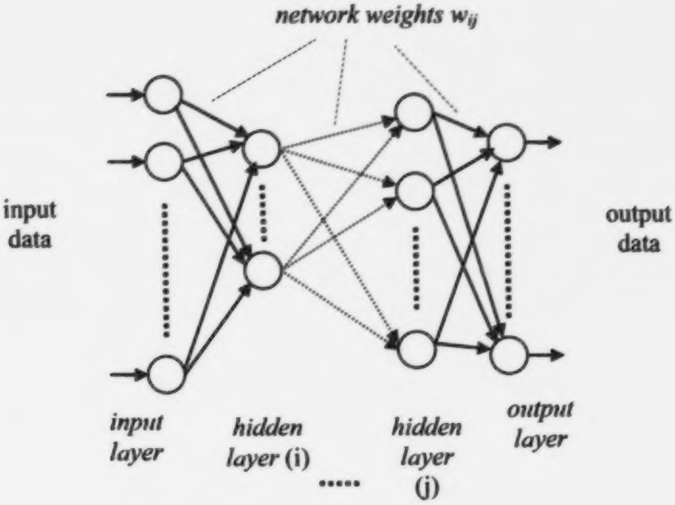


Figure 8. Artificial Neural Networks Schematized

The definition of the *output layer data* implies the choice of the desired type of forecasts and their numerical representation. According to the needs identified in the simulation model requirements, the choice retained (that could be easily changed with respect to newly identified requirements) is to make use of the forecasted information in terms of released number of containers daily, by type:

- First output: the number of ramped 20 feet containers leaving the port on day j;
- Second output: the number of ramped 40 feet containers leaving the port on day j.

The choice of day j for which the forecasts are provided (with respect to the current information available) depends on the forecasting scheme used. A *zero-days delayed prediction scheme* corresponds to forecasting volumes for the same day as the current day for which the latest information is available (inputs for day j, outputs for day j). An *f-days delayed prediction scheme* corresponds to forecasting volumes for the following f^{th} day using the current day information (inputs for day j, outputs for day $j+f$).

For the *network structure definition*, different parameters needed calibration: type of network (feed-forward, recurrent etc.), number of hidden layers, number of nodes at each hidden layer, weights initialisation, activation function, etc.

The supervised *training algorithm* had to be properly chosen as well, which implied the tuning of several other parameters: learning rate, number of training epochs, error measure etc. A thorough study was required and the results are presented in the next section.

Neural network model implementation, validation and results

For the implementation of the Neural Network-based prediction model, we used the Java Neural Network Simulator, an open source neural network simulator based on the Stuttgart Neural Network Simulator (SNNS - a simulator for neural networks developed at the Institute for Parallel and Distributed High Performance Systems at the University of Stuttgart since 1989).

During the implementation phase, comparisons between different neural network types and training algorithms were carried out in order to select the most appropriate (e.g., best fit) for the problem at hand. The neural network type retained is of the feed-forward type. Several learning algorithms were compared:

- Standard backpropagation;
- Batch backpropagation;
- Backpropagation with momentum;
- Resilient backpropagation.

The best results with respect to the speed of convergence for the learning pattern sets and for the generalisation capabilities were obtained with the *resilient backpropagation* algorithm (schematized in Figure 9). This algorithm is used for feed-forward networks and it is an *adaptive learning algorithm* which avoids the problem of "blurred" adaptation process, by changing the size of weight-updates Δw_{ij} without considering the size of the partial derivatives, but only their sign (positive or negative).

With respect to the resilient backpropagation algorithm parameter calibration, after extensive trials, the best set was found to be $\Delta_0=0.1$, $\Delta_{max}=30.0$, and $\alpha=3$. The number of neurons in the hidden layer was set at 30, as the value for which the best generalisation results were obtained (experiments with values in the [10..50] range).

Figures 9, 10, and 11 display results obtained with the feed-forward network with 13 inputs, 30 hidden neurons and 2 outputs, using the resilient backpropagation algorithm. Each figure displays part of the neural network structure in the left-top quadrant. The figures illustrate that all neurons of a layer are connected to all the neurons of the following layer. At the beginning of the learning process, all connection weight values are randomly chosen between zero and one.

The Control Panel window in the right-top quadrant contains information about the learning algorithm used (resilient propagation) and the associated learning parameters. The left-low quadrant is a log of the training iterations. It displays the numerical values of the *Mean Square Error* (*MSE*) after 500 consecutive iterations for the training and validation sets of patterns, respectively. The MSE is computed as the relative difference between the predicted values and the actual (known) values corresponding to the input data. Those numerical values are represented on the associated error graphs.

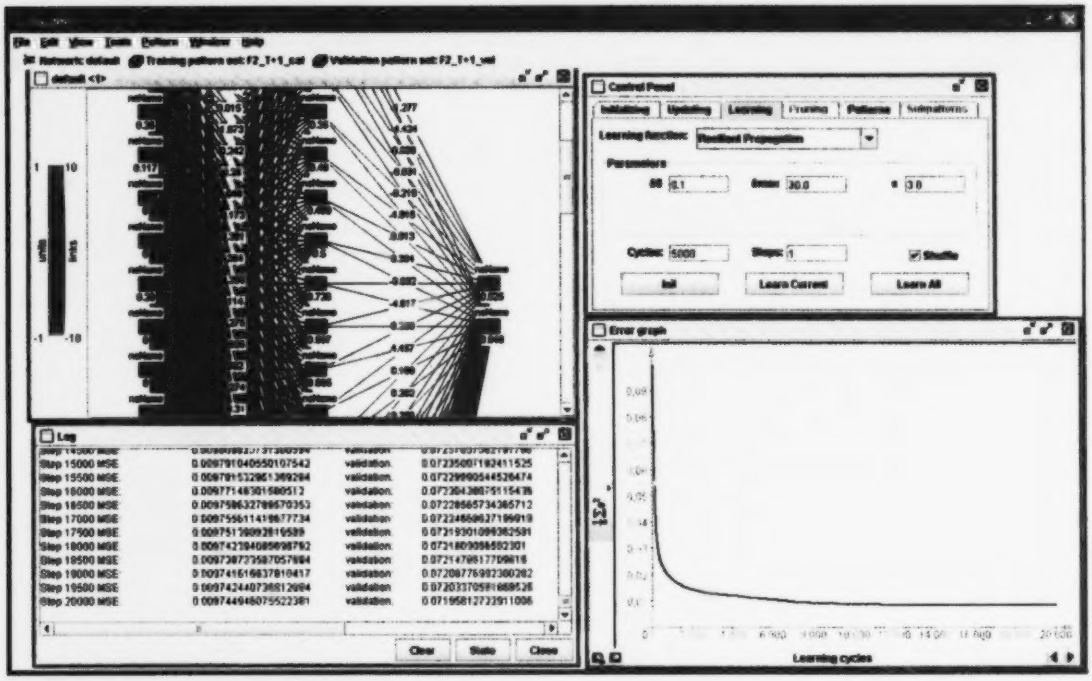
```
For all weights and biases{
    if ($\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) > 0$) then {
        $\Delta_{ij}(t) = \text{minimum}\ (\Delta_{ij}(t-1) * \eta^{+}, \Delta_{max})$
        $\Delta w_{ij}(t) = -\ \text{sign}\ (\frac{\partial E}{\partial w_{ij}}(t)) * \Delta_{ij}(t)$
        $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$
    }
    else if ($\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) < 0$) then {
        $\Delta_{ij}(t) = \text{maximum}\ (\Delta_{ij}(t-1) * \eta^{-}, \Delta_{min})$
        $w_{ij}(t+1) = w_{ij}(t) - \Delta w_{ij}(t-1)$
        $\frac{\partial E}{\partial w_{ij}}(t) = 0$
    }
    else if ($\frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) = 0$) then {
        $\Delta w_{ij}(t) = -\ \text{sign}\ (\frac{\partial E}{\partial w_{ij}}(t)) * \Delta_{ij}(t)$
        $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$
    }
}
```

Where:

$$\Delta_{ij}(t) = \begin{cases} \eta^{+}\Delta_{ij}(\mathrm{t}-1) & ,\text{if } \frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) > 0 \\ \eta^{-}\Delta_{ij}(\mathrm{t}-1) & ,\text{if } \frac{\partial E}{\partial w_{ij}}(t-1) * \frac{\partial E}{\partial w_{ij}}(t) < 0 \\ \Delta_{ij}(\mathrm{t}-1) & ,\text{else} \end{cases}$$

and where $0 < \eta^{-} < 1 < \eta^{+}$.

Figure 9. The Resilient Backpropagation Algorithm (Riedmiller and Braun, 1993)


File Edit View Tools Pattern Window Help
default <1>
Control Panel
Initializing
Updating
Learning
Pruning
Patterns
Subpatterns
Init
Learn Current
Learn All
Error graph
Learning cycles
Log
Clear
Status
Close

Network: default · Training pattern set: F2_T+1_cal · Validation pattern set: F2_T+1_val

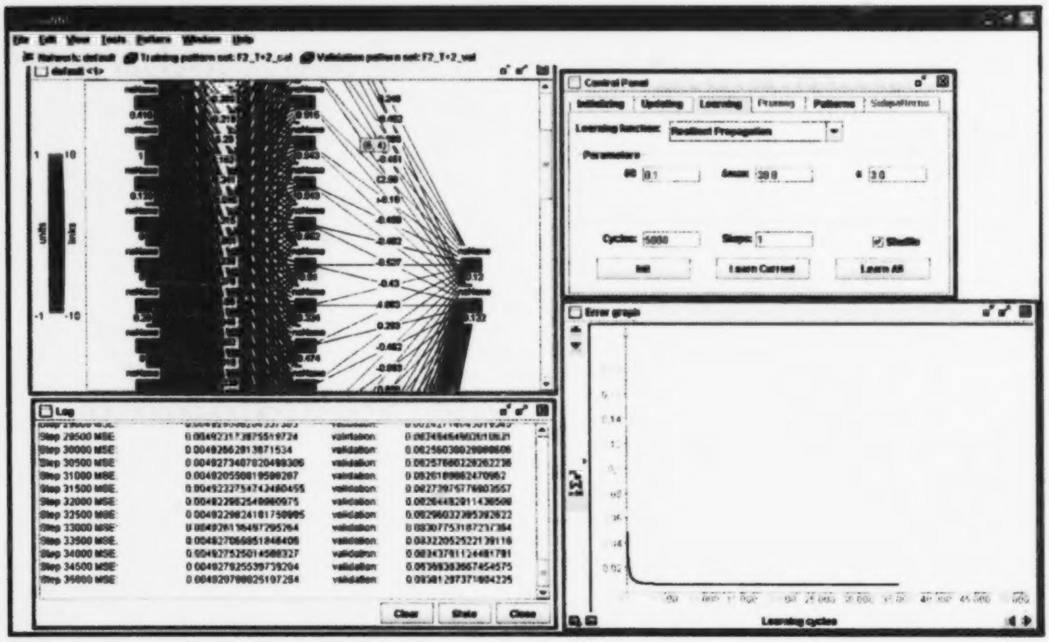
Learning function: Resilient Propagation

Parameters: δ0 0.1 · δmax 30.0 · α 3.0

Cycles: 5000 · Steps: 1 · ☑ Shuffle

```
Step 15000 MSE:    0.009791040550107542    validation:    0.07235007182411525
Step 15500 MSE:    0.009781532961369294    validation:    0.07229990544526474
Step 16000 MSE:    0.00977148301580512     validation:    0.07230438075115436
Step 16500 MSE:    0.009758632789670353    validation:    0.07228565734365712
Step 17000 MSE:    0.009755611419677734    validation:    0.07224658627195019
Step 17500 MSE:    0.009751798092819589    validation:    0.07219301099382581
Step 18000 MSE:    0.009742194085698792    validation:    0.07218093058502301
Step 18500 MSE:    0.009730723587057984    validation:    0.0721479817709818
Step 19000 MSE:    0.009741616637810417    validation:    0.07208776992300282
Step 19500 MSE:    0.009742440738812084    validation:    0.07203370591668526
Step 20000 MSE:    0.009744948075522381    validation:    0.07195812722911006
```

Figure 10. Results for a *One-day Delayed* Prediction Scheme


File Edit View Tools Pattern Window Help
default <1>
Control Panel
Initializing
Updating
Learning
Pruning
Patterns
Subpatterns
Init
Learn Current
Learn All
Error graph
Learning cycles
Log
Clear
Status
Close

Network: default · Training pattern set: F2_T+2_cal · Validation pattern set: F2_T+2_val

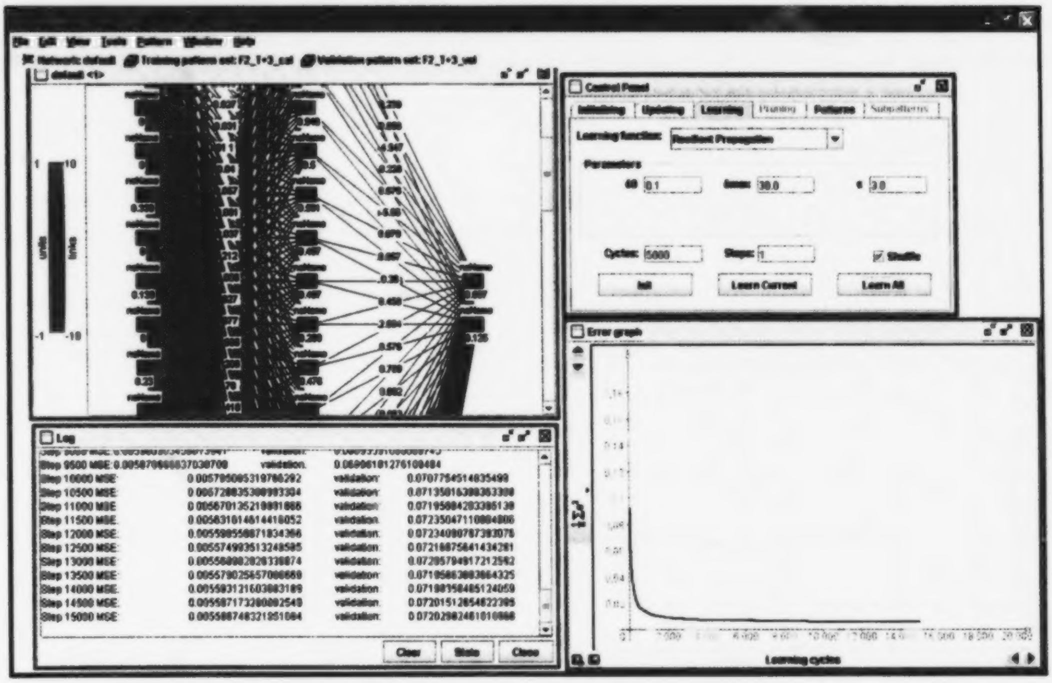
Learning function: Resilient Propagation

Parameters: δ0 0.1 · δmax 30.0 · α 3.0

Cycles: 5000 · Steps: 1 · ☑ Shuffle

```
Step 29500 MSE:    0.004923173875519724    validation:    0.08249464902610821
Step 30000 MSE:    0.004928662813871534    validation:    0.08256030029008606
Step 30500 MSE:    0.004927340782049306    validation:    0.08257860228262236
Step 31000 MSE:    0.004920550819598287    validation:    0.08261899862470952
Step 31500 MSE:    0.004923227547424480455 validation:    0.08273976776003557
Step 32000 MSE:    0.004922982540980975    validation:    0.08284482011436509
Step 32500 MSE:    0.004922982418175085    validation:    0.08296033395392622
Step 33000 MSE:    0.004926113649729576 4  validation:    0.08307753187237384
Step 33500 MSE:    0.004927066851846408    validation:    0.08322052522139116
Step 34000 MSE:    0.004927525014588327    validation:    0.08343781124481781
Step 34500 MSE:    0.004927925539739204    validation:    0.083598383667454575
Step 35000 MSE:    0.004920790025107284    validation:    0.08381287371804225
```

Figure 11. Results for a *Two-day Delayed* Prediction Scheme



```
Step 9500 MSE: 0.005870666837030700     validation:  0.06906101276100484
Step 10000 MSE:  0.005795005319786292   validation:  0.0707754514835499
Step 10500 MSE:  0.005728835300993304   validation:  0.07135016300363308
Step 11000 MSE:  0.005670135219891868   validation:  0.07195084203308138
Step 11500 MSE:  0.005631014814418052   validation:  0.07235047110084806
Step 12000 MSE:  0.005598055887183436   validation:  0.07234080787393078
Step 12500 MSE:  0.005574993513248585   validation:  0.07216875641434281
Step 13000 MSE:  0.005568982026338874   validation:  0.07205794917212592
Step 13500 MSE:  0.005579025657008680   validation:  0.07195863003864325
Step 14000 MSE:  0.005583121603883189   validation:  0.07198358485124059
Step 14500 MSE:  0.005587173280082540   validation:  0.07201512854822385
Step 15000 MSE:  0.005588748321851084   validation:  0.07202982461010866
```

Figure 12. Results for a *Three-day Delayed* Prediction Scheme

Indeed, the right-low quadrant displays graphically the information relative to the accurateness and predictive power of the models. Two curves are displayed representing the behaviour of the algorithm measured by the Mean Square Error (MSE) computed during its training phase (the black curve) and on "new", validation data (the red curve), respectively. This last is a uniformly distributed random selection of data from the original set of patterns obtained from the collected data (during the period from January 1 to May 31, 2005 for the Port of Montreal). The validation data is different from the information used to calibrate (learning phase) the neural network. The number of validation patterns is about 25 entries, whereas the number of training entries in the learning data set is about 80. This is a classical proportion with respect to neural networks training and validation methodology.

Very good learning performances could be observed in all the situations analyzed. In all three cases, the black curve stabilizes at 0.01 or lower. This shows the capacity of the neural network to "model" the transfer process of containers in the port terminal. With respect to the generalisation (prediction) capabilities, globally, all the validation errors obtained (red curves) were below 0.05, corresponding to individual values of 15%, which constitutes a very good performance.

Utilisation

In the experiments presented, and illustrated in Figures 10, 11, and 12, the number of cycles was set to 20,000, in order to give a complete image of the behaviour of the learning and validation error curves. This induces an overtraining phase, however, phase which can be noticed on the graphs when the validation curve starts to increase again after having reached its minimum. A shorter cycle would be used in practice.

In the three examples given, one, two, and three days advance prediction were considered, respectively. The training and the validation data sets were built such that the outputs of the network are the quantities of 20 and 40 feet ramped containers leaving the port one, two, or three days after the ship arrival (if ship arrival date is day j, the outputs are predicted for day $j+1$, $j+2$, or $j+3$).

The prediction mechanism works as follows: given the current day information (ship arrival, quantities of containers for particular destinations, historical arrival quantities etc), the number of 20 and 40 feet ramped containers that would leave the port the following (one, two, or three) day(s) is predicted. In our study, the prediction capabilities of the neural network model were validated using data from the port of Montreal for the months from January to May 200). The same model, i.e., the same trained neural network, should give similar good results when used with new data sets for the same terminal and the same ranges of the explanatory variables used.

If the ranges of those variables are subject to change, due to new port terminal organisation, new vessels arriving in port, new ranges of container quantities observed, and so on, the model can be dynamically updated. Using newly registered data, new learning and validation data sets must be built and used to train the neural network. It is noteworthy that this means that one does not need to change neither the structure of the neural network, nor that of the learning algorithm and parameters.

The learning process takes only a few minutes on a medium-powerful computer. Thus, generating new data sets for use with the neural network could be performed automatically and the model could thus be "updated", or retrained, daily. Such a dynamic adjustment could be implemented as long as the behaviour of the system is reasonably stable. When, on the other hand, important structural changes in the process are observed (i.e., the explanatory variables would change), a thorough calibration phase should be performed with newly collected data.

Conclusions

The neural network forecasting method was designed and implemented using historical data from CN giving arrival dates of ships in port and their respective total import volumes (data ***D*** in Figure 7), as well as quantities of containers ramped on rail cars departing daily from the port terminal. The former is taken as *input data*, while the latter forms the *output data*. An important phase of data collection and pre-processing was necessary before the designed neural network could be calibrated, trained, and the generalisation (prediction) capabilities assessed.

The input and corresponding output data were used to build training and validation pattern sets for the proposed ANN model. These were fed to the network and, after a sufficient number of learning epochs, the "trained" neural network was obtained. A different data set, with the same characteristics, was used for validation purposes and the predictive performances of the network were analysed. Thus, the network trained and validated (the model obtained) can be used to predict expected quantities and types of released containers, on a daily basis, available for loading on CN trains. The predictions can be made for the current day or for the following days (one, two, three days in advance). Such (very) short-term forecasts present the advantage of being obtained using the *latest information available in the system* (all the information available at the moment of decision making).

Very good learning performances could be observed in all the situations analyzed. This shows the capacity of the neural network to "model" the transfer process of containers in the port terminal. Similarly good performances were observed with respect to the prediction capabilities as well. Studies are under way to perform further tests concerning the relevance and the dimension reduction of the number of entries (variables) for the prediction model. It is useful to recall that, given the high complexity and nonlinearity of the studied process as well as the dynamic utilisation requirements, an analytical model would not constitute a valid demand prediction model.

All the data we assumed necessary at the beginning of the study for prediction purposes were in fact not available. The incompleteness of the data had to be taken into consideration. Incomplete data patterns were eliminated from the learning and validation data sets, as well as data supposed to be erroneous (e.g., working days where the quantities of ramped containers quantities were zero in the original data files). Further calibration and utilisation of the prediction model should be based on a more comprehensive set of data.

The neural network model was built to predict the quantities of containers *already ramped* (at the exit of the port), because this was the data available at the time of the study. This does not constitute, from our point of view, the *real demand* for the CN intermodal services. If new data would become available (e.g., relative to the released containers at port exit – not necessarily ramped and departed on train containers), the neural network could be easily trained again to take these new sets of data into account and thus predict information closer to the *real demand* for the intermodal transportation services.

Quality and quantity of data are widely recognized as important issues in the development of neural network models (Lyons *et al.* 1998). Using real data collected from the field, instead of simulated data, makes our research even more difficult and challenging, while also enhancing the importance of the results obtained.

VII. Simulation

This work dimension develops the simulation model and software required for the validation and evaluation of the models, concepts, and procedures proposed in the other work directions. It is worth remembering, however, that 1) the simulator is not a deliverable to be transferred to a partner; it is rather a tool to be used by the CRT team; and 2) the goal is to develop a first generation of the simulator to be refined and enhanced in subsequent projects.

We used the simulation software ARENA (Version 8.1, Kelton *et al.* 2004) for this development. ARENA is a well-known system that offers several advantages including an easy-to-use graphical interface for model building that is useful both for the initial development of a simple model and the subsequent refinements, a complete and powerful simulation language, and a comprehensive set of probability and statistics features for model specification and result gathering. Several ARENA licences are installed on the CRT computing network.

Following extensive exchanges with the CN group, the *planning and requirements* task yielded the specification of a generic simulation model of rail intermodal transportation, including the network, terminals, services, demand, etc., with possible attributes and performance measures. The specification was sufficiently general to allow for the simulation of different policies and scenarios as well as to permit further enhancement and refinement of the simulation models. The generic simulation model was then applied to the specific case of the Eastern CN network composed of the four main intermodal terminals: Chicago, Toronto, Montreal, and Halifax. Their connections with the rest of the railway network were also represented in the model. The simulation takes into account the rail lines between the terminals, the trains (two types of cars and the power engines), the demand (loaded and empty 20' and 40' containers), the services and transportation plan (train length and composition, block length), and main CN operation rules and policies in order to mimic the services offered by CN between these origins and destinations for container transportation.

A number of simplifying hypotheses, decided jointly with the CN team, were used to build this first application:

- Associate delays to operations, instead of trying to get to a very low level of detail;
- Avoid line utilisation details; To simulate incidents, for example, a different, longer, station-to-station transit time is to be generated;
- Do not detail classification yards, use only aggregated capacity figures when required; In fact, sufficient handling capacity in yards was assumed;
- Include empty container movements; Empty containers are used to "fill up" blocks when the quantity of loaded cars is lower than the block capacity;
- Assume cars and power engines are always available in the system, but use the *Base Plan* to keep inventories of cars;
- Use averages on loading operations (with respect to composition of blocks, contents, etc.);
- All cars are assumed to offer 2 slots (2 FEUs) and double-stack capability. The car types considered are 1) long and heavy, 2) long and light;
- Two types of containers are to be considered, 20 feet and 40 feet (the unit measure to be used is the slot = length in feet/35);

A schematic description of the generic simulation model is given in Figures 13 to 19.

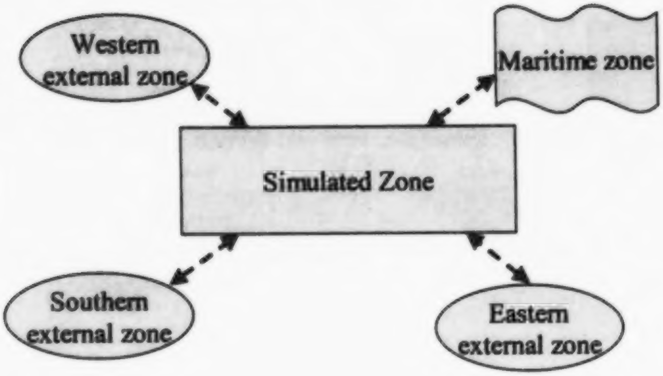

Western external zone
Maritime zone
Simulated Zone
Southern external zone
Eastern external zone

Figure 13. Simulation Model: General View

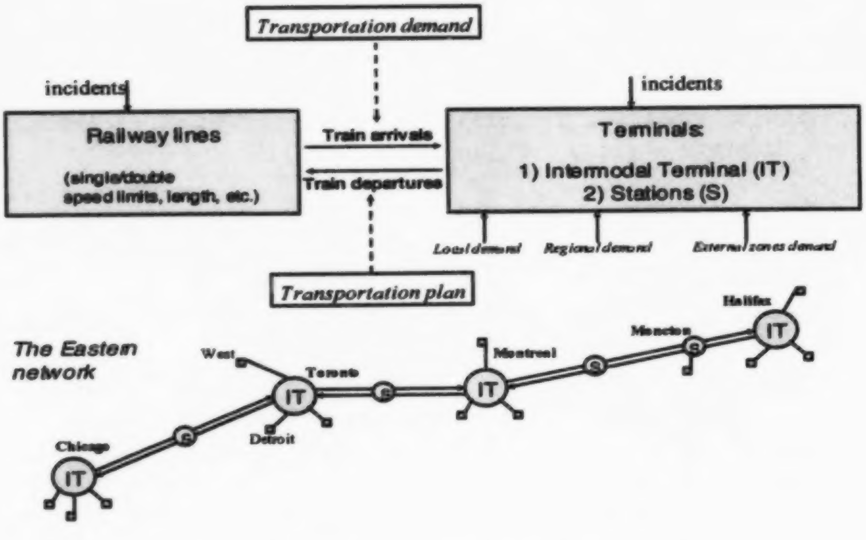

Transportation demand
incidents
incidents
Railway lines
(single/double speed limits, length, etc.)
Train arrivals
Train departures
Terminals:
1) Intermodal Terminal (IT)
2) Stations (S)
Local demand
Regional demand
External zones demand
Transportation plan
The Eastern network
West
Toronto
Detroit
Chicago
Montreal
Moncton
Halifax
IT
IT
IT
IT
S
S
S
S

Figure 14. The Simulated Eastern Network

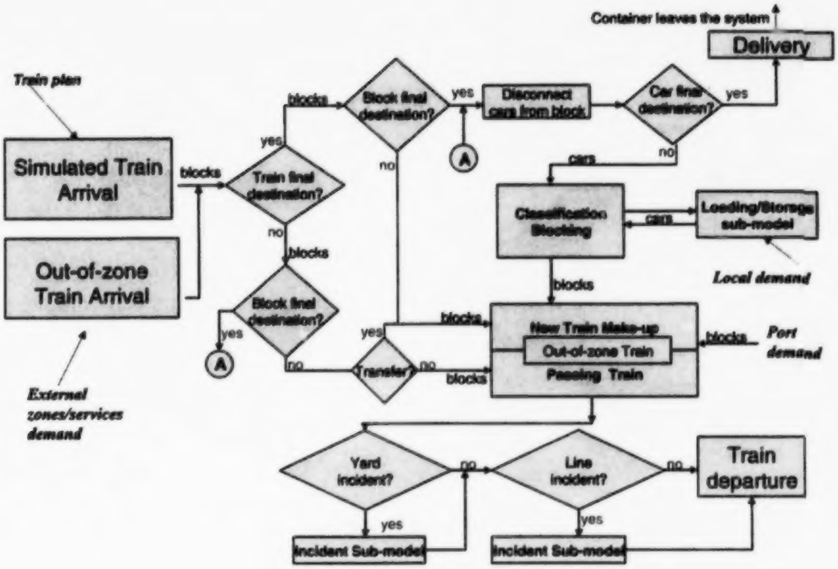

Container leaves the system
Delivery
Train plan
blocks
Block final destination?
yes
Disconnect cars from block
Car final destination?
yes
yes
A
no
no
cars
Simulated Train Arrival
blocks
Train final destination?
Classification Blocking
cars
Loading/Storage sub-model
no
blocks
Local demand
Out-of-zone Train Arrival
blocks
Block final destination?
blocks
yes
New Train Make-up
blocks
Port demand
yes
Out-of-zone Train
A
no
Transfer?
no
blocks
Passing Train
External zones/services demand
Yard incident?
no
Line incident?
no
Train departure
yes
yes
Incident Sub-model
Incident Sub-model

Figure 15. Intermodal Terminal Sub-model

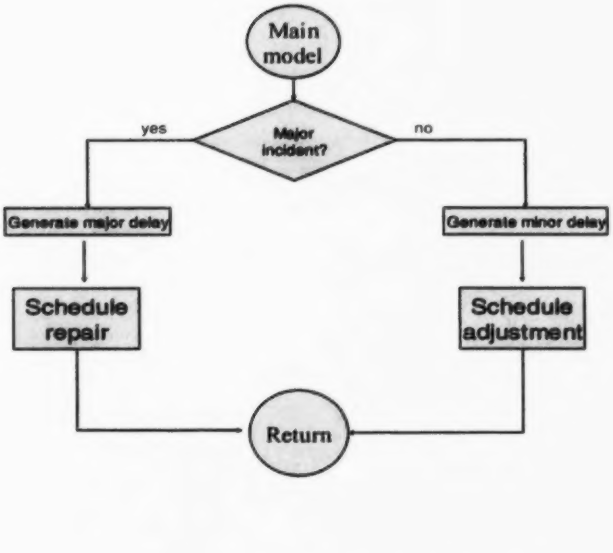

Main model
yes
Major incident?
no
Generate major delay
Generate minor delay
Schedule repair
Schedule adjustment
Return

Figure 16. Incident Sub-model

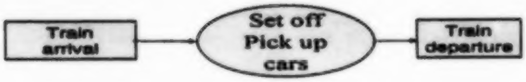

Train arrival
Set off Pick up cars
Train departure

Figure 17. Station Sub-model

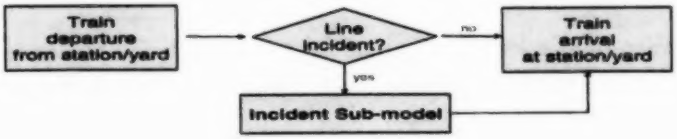

Train departure from station/yard
Line incident?
no
yes
Train arrival at station/yard
Incident Sub-model

Figure 18. Line Segment Sub-model

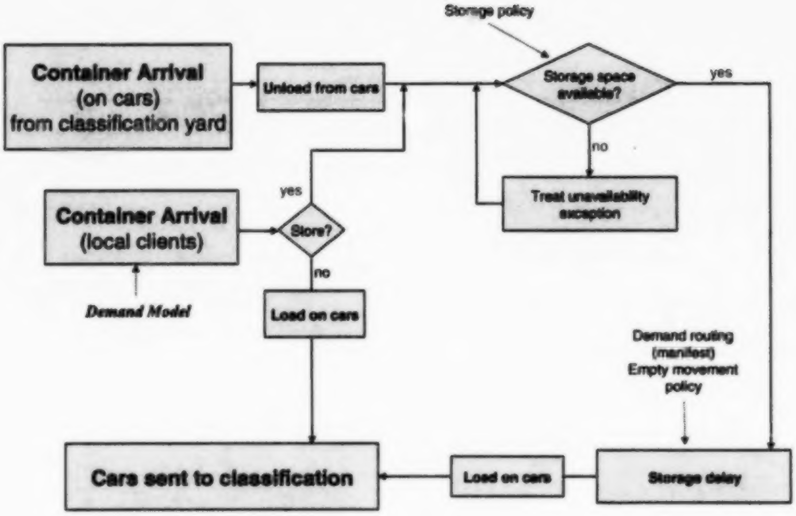

Storage policy
Container Arrival
(on cars)
from classification yard
Unload from cars
Storage space available?
yes
no
Treat unavailability exception
yes
Container Arrival
(local clients)
Store?
no
Demand Model
Load on cars
Demand routing
(manifest)
Empty movement
policy
Cars sent to classification
Load on cars
Storage delay

Figure 19. Container Storage at Yard Sub-model

The validation of the simulation model has been performed by comparing the statistical performance of the results obtained by the simulation on actual data to the historical performance of the current system (same "data"). More precisely, to validate the simulator, we performed five executions of the simulation over five different demand files that we generated randomly according to the statistics estimated from the real demand the CN provided us (Section V). Each run simulated 90 days of operations. For each run, we recorded for each origin-destination combination the mean and the standard deviation of the delivery time in days, as well as the number of containers that completed the corresponding movement during these 90 days.

A thorough comparison of the values observed in the various runs indicates that the simulator is very stable: the five repetitions yield very similar statistical results for the 90 days of operations. The results produced by the simulation model, including the random demand generation process, also fit very well the actual data provided by CN. For all active O-D pairs (i.e., pairs with positive demand in the received data), the transit time obtained by the simulation is the same or very close to actual figures. A significant discrepancy was observed for only two O-D pairs. Both O-D pairs relate to the fact that three demands compete for space on a very short block of one train service. This discrepancy was sorted out following further discussions with CN personnel: it turns out, in fact, that the waybill that had been implemented in the simulator for traffic from one of the problematic O-D pairs was incorrect: it involved sending containers through the wrong intermediate destination. It is interesting to note how the results of the simulation point out to errors in data. Verifications are now underway with CN to validate all the waybills used in the simulator; corrected waybills (besides the ones already identified, there should be very few modifications) will be implemented shortly.

Globally, the results of this validation step indicate that this first generation simulator may be used to analyze various scenarios for the CN Eastern intermodal service network.

VIII. Intermodal Service Concepts

This work dimension addresses the supply side of the intermodal transportation system and is thus at the core of the project goals. The objective is to explore – define and analyze – possible service and operations models that integrate international intermodal traffic. It is important to recall that the system proposed by CN is truly innovative as there are few other such systems around the world that we are aware of and none on a scale and scope envisioned here. This work dimension is therefore the most exploratory of the project.

Understanding the CN intermodal operations

A key element in this part of this project was gaining a clearer understanding of CN operations, as well as acquiring the corresponding data. This work was performed mainly during the first half of the project through a number of meetings with CN personnel, as well as other key actors in Halifax and Montreal, as was already described in the previous sections. These meetings allowed us to put together a description of the current operation rules and policies of the service, including the specification of the goals and objectives that lead to their introduction. We were also able to collect some information on customer service and links and exchanges with partners and customers, as well as on current information and decision support systems. Current operation rules and policies are listed in Section V. With respect to data, the data relevant to this work dimension was procured under the data collection activities described in the previous sections.

Identification of key issues

Our exchanges with CN personnel and port officials in Halifax and Montreal allowed us to sharpen significantly our vision of the real issues involved in the improvement of the intermodal service with respect to the maritime import traffic. A cornerstone in the new *modus operandi* of CN is the objective of achieving (near-) "full asset utilization". Scheduled services and advanced bookings are important in this context because they provide for regular and predictable operations, in which the most critical assets can be turned out around efficiently. It became quite clear that any contemplated change in operating rules should respect this guiding principle.

The booking system used in the land terminals of CN for its intermodal traffic cannot be extended to the import operations in a simplistic fashion (e.g., by allowing shippers or freight forwarders to book railcars for their containers), because the creation of railcar blocks for trains services operating from Halifax or the Port of Montreal is not under the direct control of CN, but rather of port terminal operators. On the other hand, CN has control on the supply of empty cars that are provided to the ports and on the overall length of the trains that operate from the ports.

While regular operations are desirable from the CN standpoint, as explained above, they tend to be disrupted by a number of phenomena, including the fact that arrivals of ships in ports are not regularly distributed. When this occurs, CN operations out of the corresponding port are severely strained: there might be several days without arrivals, followed by a large turnout of arriving containers. With fixed, regular service, these peaks in demand are not easily accommodated: either exceptional *ad hoc* measures must be implemented or huge backlogs of arriving containers pile up in ports.

From the previous discussion, it follows that ***flexibility*** to adjust to fluctuations in demand is probably the most critical issue at this time. But not flexibility provided without regards to cost and other operational issues. In fact, the central point of our proposals for improvement is to try to achieve flexibility without foregoing efficiency, i.e., full asset utilization. Furthermore, our goal is not to provide flexibility after the fact, curing as well as one can crises, but rather in a preventive mode using as much as possible advance information on demand, which could be derived from instance from the container release time model and other information available from shipping lines.

An optimization-based decision support tool

A major stumbling block to the introduction of more flexibility to respond to fluctuations in demand is the fact that any contemplated change in train composition will have downstream effects that may undo any short-term benefit from this change. In fact, when any significant change to train service or block length is examined, it is extremely important to determine how "normal" operations will be resumed at some point in the future (usually, the sooner, the better). Unfortunately, train operations are too complex to allow many alternatives to be considered at any given point in time, if traditional planning and management techniques are used. We propose to circumvent this difficulty by resorting instead to a decision support tool based on a *network optimization model*. The purpose of this model is to suggest to CN managers "optimal" block lengths and number of empty cars moved, and resulting train lengths, for all services within the network considered, over a planning horizon long enough (three to four weeks) to fully capture the effects of possibly any change and to allow enforcing a return to normal (planned) operations. A rough (non-mathematical) description of this model follows. A full mathematical description is provided in Annex A

For each train service under consideration, the model considers as decision variables the number of loaded cars for each destination in each possible block, as well as the number of empty cars carried over each service leg. Simple conservation of flow constraints guarantee that loaded and empty cars flow correctly through time over the intermodal network. Other decision variables allow the model to choose, among feasible options, the time at which the containers released on a given day in a given port for a given destination will be "ramped" and included in a block of a departing train. The main inputs of the models are:

1. Data on car flows relative to current train movements;
2. Planned values for block and train lengths;
3. Expected demands (expected number of containers to be released or ramped at each port/yard over some *prediction horizon*).

It is important to note that while we expect to have perfect information on current flows, we assume that demand predictions are available only for a few days ahead (normally, these predictions should be derived from information provided by the container-release time model for ports and by advanced bookings (supplemented by relevant statistical information) for inland yards. These few days ahead constitute the *prediction horizon*. Past this horizon, we use for model purposes long-term statistical demand estimates.

In its current version, the optimization model imposes that all the containers that should arrive to their destination during the planning horizon do and minimizes departures from planned values for train length and composition (i.e., the objective function of the model is simply the sum of all absolute deviations from planned values). The model could be refined to allow late arrivals of containers at some cost, to penalize more heavily deviations from the plan late in the planning horizon, or to include additional constraints on block or train make-up. Adding any of these refinements would not present any special challenge since we plan to solve the model using powerful commercial optimization software, namely CPLEX.

It should be stressed that this model, if it were to be used to support actual operations, should be implemented in a *rolling horizon* mode, i.e., only the decisions pertaining to the first period(s) it covers should be implemented. Data should be re-actualized and the model rerun to determine the proper decisions to implement in later periods.

Model assessment

To assess the usefulness of the model, we have considered three scenarios. The first one corresponds to fairly normal operations with average fluctuations in demand for each origin-destination pair. The second corresponds to a situation encountered in Halifax where sometimes ship arrivals concentrate on the later days of the week, instead of being evenly spread out over the week: for each Port of Halifax origin (i.e., HHL or FCV), average demands for each destination are multiplied by 150% on Fridays and Saturdays, and by 50% on Mondays and Tuesdays. The third scenario looks at extreme variations stemming from severe storms on the North Atlantic. Under this scenario, there is a week of storms where demands originating from the Port of Halifax are only 25% of their usual values, while other demands are unchanged. During the following week, delayed ships arrive, raising demands to 175% of their usual values. It is to be noted that all three scenarios are run for a full season (90 days).

For each of these scenarios, we consider two extreme situations with regards to demand estimation. The first one is the *uninformed case* under which all demand estimates are simply set to their historical average (i.e., there is no attempt to take advantage of real-time data); in this case, the model functions essentially in reactive mode and we can determine how using the model can help in solving crises. The second situation is the *omniscient case* in which we feed the actual demand values that will be used by the simulation to the model. This situation is clearly an ideal one, since we cannot expect any prediction method to perform as effectively. For the two latter scenarios, we may also consider an intermediate situation, the *generally informed case*, in which the model is not provided with actual demand values, but with appropriate averages, given the scenario (i.e., in the second scenario, one knows that ships arrive late in the week and that demand fluctuates accordingly). These information situations will allow us to better bound the improvements that the model could provide.

Experimental results

Simulation runs were first performed for the base case of the three scenarios described above, i.e., the situation where the network optimization model is not used. This base case can serve as the reference point against which results obtained when the network model is activated should be compared. Average delays (computed over 5 runs) between ramping and delivery dates for the O-D pairs originating in Halifax are reported in Table 4. These results clearly show the stability of the simulation.

Origin	Destination	Scenario 0	Scenario 1	Scenario 2
HHL	MTL	2.0	2.0	2.0
FCV	MTL	3.1	3.0	3.1
FCV	CHI	4.1	4.5	4.8
HHL	BIT	2.3	2.4	3.0
FCV	DET	2.6	2.9	2.6
FCV	BIT	2.4	2.7	2.8

Table 4. Average Delays for each Scenario for the O-D Pairs with a Port of Halifax Origin

Further runs focussed on 6-week simulations of scenario 0, in which the network model was invoked after 3 weeks of simulation to adjust the block lengths afterwards. The results display modifications to the length of nine blocks. This adjustment resulted, for the operations with optimization, in a significant decrease (of the order of 39% and 61%, respectively) in the average delays of the two problematic O-D pairs (see Section VI) compared to the scenario without optimization.

It should be pointed out that these runs where performed under the omniscient information scenario to see how much improvement the model could provide under ideal conditions. Additional runs involving a 30-day horizon in which the network model was again invoked after 21 days were also performed and produced similar results.

Analysis and discussion

It is interesting to note that moving from a stable demand (scenario 0) to one with cyclic variations (scenario 1) has a significant, albeit small, impact on delays for most of the affected O-D combinations. In those cases where the impact is negligible, we conjecture that the overall demand is low when compared to planned block lengths. The impact of more radical variations in demand patterns (scenario 3) is, as expected, more important. We expect, however, that the use of the network model will help mitigate these negative impacts.

The results obtained when using the optimization model jointly with the simulator clearly show that such a tool may be very useful in identifying problem spots. In our case, the result that some block lengths are inadequate when one uses the waybills currently implemented by the simulator, lead to the identification of two problems. The first was simply that some of the routings used in the simulator did not correspond to the ones actually operated by CN. The second emphasized the network effect of service design and operations: the initial discussion with CN personnel concluded that some not fully "intermodal" services did not need to be included. The results point to the fact that such services should be included in order to achieve a correct image of actual operations.

The experimental results also indicate that the optimization model is capable of proposing reasonable solutions to tackle imbalances in traffic. This last remark points out to the proper way of looking at the optimization model: this is a *decision-support tool* intended to provide decision-makers with suggestions about courses of action, not a substitute to decision-makers. It should also be stressed that, in normal operations, the model would be fed with demand data provided by the neural network model for ports and by advance bookings for inland terminals, and would be used in a rolling-horizon mode.

A last point of concern regarding the experimental runs involving the simulator and the optimization model is that these are not only very time-consuming, but also unstable, in the sense that strange software interactions occur when one tries to invoke directly the optimization software from the simulator. In fact, there seem to be severe integration problems between the simulation and the optimization environments, in spite of claims to the contrary. A rigorous assessment of the software involved in these tests will therefore need to be conducted before making any further developments.

XI. Conclusions and Perspectives

This section closes the Final Report. It summarizes the findings and results of the project, discusses challenges ahead, and proposes a roadmap to develop the methods and instruments required to address them.

The project achieved its goals. We analyzed the operations of a new and innovative intermodal rail service, which operates under a full-asset-utilization policy and implements a reservation system for all operations, except import from deep-sea navigation. We also achieved a level of understanding of container port terminals appropriate for the scope and means of the project. The analysis went further than the description of the operations and encompassed information exchanges among the major stakeholders involved. We have also developed prototype tools for the planning and analysis of operations.

The meetings with port and container port representatives, CN personnel, and other stakeholders emphasized the fact that the development of better communication and cooperation among stakeholders is critical in order to improve the efficiency and profitability of the services offered. This should be possible if the adequate tools for doing so are deployed and used. The continuing evolution of mentalities and corporate culture is also an essential factor in this respect.

A number of critical points concerning the informational and decisional flows related to intermodal traffic emerged from these meetings and discussions. What follows illustrates our current understanding of some key issues. We want to emphasize that these are preliminary impressions that require further data and analysis. This topic warrants, in fact, a project all by itself. The general picture that starts to emerge could form the starting point of this study:

- Much information exists and is exchanged between the principal stakeholders: maritime shipping lines, port authorities, port terminal management, Canada Customs and Agricultural control authorities, freight forwarders and trucking firms. Regular meetings bring together, in person or virtually, several of these entities to exchange information and update plans.
- Most of these information transmissions seem to be strictly unidirectional (e.g., shipping line to Customs and shipping line Port or to rail carrier). Moreover, the "same"

information may be sent to different organizations at different times. This results in multiplication of efforts and hampers coordination of various stakeholder activities. In fact, the information is used for very specific and local purposes, sometimes being registered in data bases, sometimes being discarded after use. It would also appear that some deficiencies exist at the level of (quasi) real time data (e.g., availability of containers following Customs inspection). A better sharing and utilisation of this information in an integrated intermodal transportation management system that would include "all" stakeholders presents a high potential to support intelligent decision making and maximize the effectiveness and competitiveness of the participating stakeholders.

- It was sometimes difficult to obtain answers to specific questions, either because the interlocutor was not entitled to disclose that kind of information or because several answers were possible for a given question, as several actors have sometimes different views with respect to the same issue.
- Not all information exchanged is used for planning and management. An illustration of this fact is the stream of information regarding ship arrival times that did not seem to be used to manage the flow of trains and cars.

The scope of the project did not include the analysis of port container terminal operations. Based on actual observations, however, it is our strong belief that integrated optimization of operations would help improve the performance of the movements of containers in port terminals (e.g., the actual way of positioning of containers on the ship does not always take into account the information about what containers have to be unloaded at the destination ports and thus, in order to comply with the ship balance and mass center requirements, very often a large number of movements is needed for unloading containers from ships, which is time consuming and costly) and between the terminal and the sea and land transportation modes, resulting in the maximization of asset utilization and profits.

With respect to the intermodal rail operations, it was very interesting to observe that while operational decisions are taken on the basis of predefined transportation plans, updates and modifications with respect to variations in demand and system status are performed continuously, but not systematically. Moreover, a lot of "expert information" and "experience" is "hidden" within the decisional chain with respect to the "best way to react to changes in demand and environment"; This could/should be better identified, characterized, and integrated in an advanced ITS toolset for decision support of current operation and the design of appropriate plans and modes of operation to react to forecast or certain changes in the economic or regulatory environment.

We have thus developed asset of methodologies and tools. We propose a forecasting methodology that could use a broad set of data, which provides the capability to integrate newly available data. Moreover, this powerful and flexible methodology can be used in a dynamic context and thus could be integrated to a procedure that dynamically adjusts the offer of service according to demand.

We examined the issue of defining a booking system for import loads. It appears that for as long as intermodal services differentiated by delivery time, and tariff, are not widely used in international intermodal container transportation, there is not much incentive to develop a sophisticated reservation system for containers arriving by sea. One of the major issues, in fact, is the contrast between a rigid service design and schedule and the variation of demand in terms of availability and amplitude.

We have therefore addressed this issue and proposed a methodology to dynamically adjust the offer of service, given planned and forecast demand variations. The initial experiments are very encouraging and this is a line of research and development that should be continued. We expect even more interesting results when this methodology is combined to the proposed forecasting approach.

As planned, a first-generation simulation model has been developed and applied to the study of the CN Eastern intermodal service network. This development will be instrumental in analysing and evaluating future proposals for service design and operational rules.

Several avenues are open as a follow up on this project. First, there appears to be a need for a methodology to study in an integrated way ports and their land transportation networks. The ITS component should be of prime interest in this study. As in many other settings, this project should not only consider the information technology developments, but also the "intelligence" of the systems, the various decision technologies, mainly operations research-based, which transform information into decisions.

Much work is yet required on the intermodal rail side. The service adjustment and the forecasting methodologies require more development and testing. The integration of these methodologies into a comprehensive decision-support method is a companion issue of much interest and significant challenge. At longer term, the issue of the design of the services and schedules is of importance. A smoother and more profitable operation could probably be achieved by building a transportation plan that accounts more closely for the actual characteristics of demand. A proper study of demand is also required, not only on the import side, but also on the export and domestic fronts, including the response to the reservation system. The integration of the latter and the service design and adjustment methodologies presents considerable challenges but could result in a significant improvement in performance.

Bibliography

Abdelwahab W., and Sayed T., Freight mode choice models using artificial neural networks, *Civil Engineering and Environmental Systems* 16(4), 1999, pp 267-286

Al-Deek, H.M., Comparison of Two Approaches for Modeling Freight Movement at Seaports, *ASCE Journal of Computing I Civil Engineering* 15(4), October 2001a, pp. 309-319

Al-Deek, H.M., Use of Vessel Freight Data to Forecast Heavy Truck Movements at Seaports, *Journal of the Transportation Research Board* 1804, 2002, pp. 217-224

Al-Deek, H.M., Which Method is better for Developing Freight Planning Models at Seaports: Neural Networks or Multiple Regression?, *Journal of the Transportation Research Board* 1763, December 2001b, pp. 90-97

Armbruster, W., *Port Congestion Plagues Shippers*, 2004, http://www.shippingdigest.com/news/article.asp?ltype=feature&sid=564

Assad, A.A., Analytical Models in Rail Transportation: An Annotated Bibliography, *INFOR* 19, 1981, pp. 59-80

Assad, A.A., Modelling of Rail Networks: Toward a Routing/Makeup Model, *Transportation Research Part B* 14, 1980, pp. 101-114

Assad, A.A., Models for Rail Transportation. *Transportation Research Part A: Policy and Practice* 14, 1980a, pp. 205-220

Bailey, A., The Status of Car Scheduling: A Railroad Industry Review, presented at *INFORMS Rail Special Interest Group Spring Roundtable*, Los Angeles, CA, April 1, 1995

Bansal, K., Vadhavkar S., and Gupta, A., Neural networks based forecasting techniques for inventory control applications, *Data Mining and Knowledge Discovery* 2(1), 1998, pp. 97-102

Baugher, R., Norfolk Southern Integrated Transportation Management System: "Seamless" Service Management, presented at *Transportation Research Forum*, New York City meeting, 1993

Bontekoning, Y.M., Macharis, C., and Trip, J.J., Is a new applied transportation research field emerging? – A review of intermodal rail-truck freight transport literature, *Transportation Research Part A* 38, 2004, pp. 1-34

Bostel, N., and Dejax, P., Models and Algorithms for Container Allocation Problems on Trains in a Rapid Transshipment Shunting Yard, *Transportation Science* 32(4), 1998

Boyle, M.D., Business-to-Business Marketplaces for Freight Transportation, *Master of Engineering in Logistics dissertation*, Engineering Systems Division, MIT, 2000

CaLTrade Report, *Pacific Port Congestion Intensifies*, 2004, http://www.caltradereport.com

Chajakis, E.D., Management science for marine petroleum logistics, in *Decision Making: Recent Developments and Applications*, Zanakis, S.H., Doukidis, G., and Zopounidis, C. (eds.), Kluwer Academic Publishers, 2000, 169-185

Cheung, R.K. and Chen, C.-Y., A Two-Stage Stochastic Network Model and Solution Methods for the Dynamic Empty Container Allocation Problem, *Transportation Science* 32(2), 1998, pp. 142–162

Christiansen, M., Fagerholt, K., Nygreen, B., and Ronen, D., Maritime Transportation, in *Transportation, Handbooks in Operations Research and Management Science*, C. Barnhart and G. Laporte (eds.), North-Holland, Amsterdam, forthcoming, 2005

Christiansen, M., Fagerholt, K., and Ronen, D., Ship Routing and Scheduling: Status and Perspectives, *Transportation Science* 38(1), 2004, pp. 1-18

Comtois, C. and Slack, B. Le port de Montréal: un système portuaire inachevé?, *Cahiers Nantais*, No. 55/56, 2002, pp. 141-151

Containerisation International, *On the Front Line*, CI December 2003

Cordeau, J.-F., Toth, P., and Vigo, D., A survey of Optimization Models for Train Routing and Scheduling, *Transportation Science* 32(4), 1998, pp. 380-404

Crainic T.G., *Freight ITS*, Discussion Paper, WG2.1 Intermodal Freight, Pre-clearance & Logistics, ATLANTIC Canadian Project, 2004, www.crt.umontreal.ca/atlantic

Crainic, T.G. and Gendreau, M., Advanced Fleet Management Systems and Advisors: Converging Decision Technologies for ITS and E-Business, *ITS World Congress*, Madrid, 2003

Crainic, T.G. and Kim, K.H., Intermodal Transportation, in *Transportation, Handbooks in Operations Research and Management Science*, C. Barnhart and G. Laporte (eds.), North-Holland, Amsterdam, forthcoming, 2005

Crainic, T.G. and Laporte, G., Planning Models for Freight Transportation, *European Journal of Operational Research* 97(3), 1997, pp. 409–438

Crainic, T.G. and Roy, J., OR Tools for Tactical Freight Transportation Planning, *European Journal of Operational Research* 33(3), 1988, pp. 290-297

Crainic, T.G., Gendreau, M., and Dejax, P.J., Dynamic Stochastic Models for the Allocation of Empty Containers, *Operations Research* 43, 1993, pp. 102–302

Crainic, T.G., Long Haul Freight Transportation, in *Handbook of Transportation Science*, second edition, R.W. Hall (Ed.), Kluwer Academic Publishers, Norwell, Ma., 2003, pp. 451-516

Crainic, T.G., Network Design in Freight Transportation, *European Journal of Operational Research* 122(2), 2000, pp.272–288

Crainic, T.G., Rail Tactical Planning: Issues, Models and Tools, in *Freight Transport Planning and Logistics*, Bianco, L. and Bella, A. L., (eds.), Springer-Verlag, Berlin, 1988, pp. 463–509

Crainic, T.G., *Un modèle de planification tactique pour le transport ferroviaire des marchandises*, PhD thesis, Université de Montréal, Montréal, QC, Canada, 1982

Dejax, P., and Bookbinder, J.H., Goods Transportation by the French National Railway (SNCF): the Measurement and Marketing of Reliability, *Transportation Research Part A*, 25(4), 1991, pp. 219-225.

Dejax, P.J. and Crainic, T.G., A Review of Empty Flows and Fleet Management Models in Freight Transportation. *Transportation Science* 21(4), 1987, pp.227–247

Delorme, L., Roy, J., and Rousseau, J.-M., Motor-Carrier Operation Planning Models: A State of the Art, in *Freight Transport Planning and Logistics*, Bianco, L. and Bella, A. L. (eds.), Springer-Verlag, Berlin, 1988, pp. 510–545

Dharia A, and Adeli H., Neural network model for rapid forecasting of freeway link travel time, *Engineering Applications of Artificial Intelligence* 16(7-8), 2003, pp. 607-613

Dougherthy, M.S., A review of neural networks applied to transport, *Transportation Research Part C* 3(4), 1995, pp. 247-260

Dougherty, M.S., Applications of Neural Networks in Transportation, *Transportation Research Part C* 5(5), 1997, pp. 255-257

Dürr, E., and Giannopoulos, G.A., SITS: a system for uniform intermodal freight transport information exchange, *International Journal of Transport Management* 1, 2003, pp. 175-186

Eckhardt, J., Permala, A., Scholliers, J., and Leviäkangas, P., Integration of Tracking & Tracing Systems in Multimodal Transport Chains, *International ITS Congress*, Madrid, 2003.

Ermol'ev, Y.M., Krivets, T.A., and Petukhov, V.S., Planning of Shipping Empty Seaborne Containers, *Cybernetics* 12, 1976, pp. 644–646

EUROBORDER, Final Report for Publication, Contract No. WA-95-SC.153, Supported by the Commission of the European Communities – DGVII, 1998, 96p.

European Conference of Ministers of Transport, United Nations Economic Commission for Europe Statistical Division and European Union Eurostat, Glossary for Transport Statistics, second ed. See also http://www1.oecd.org/cem/online/glossaries/

European Research Program, Doc.Euret/387/94, *Economic and Technical Research of the Transfer of Goods: Design and Evaluation of Rapid Transfer*, Management summary, Brussels, June 1994

European Research Program, Doc.Euret/405/94, *Traffic Volume: Present and Future, final report*, Brussels, 1994

Faghri, A., Aneja, S., and Vaziri, M., Estimation of percentage of pass-by trips generated by a shopping center using artificial neural networks, *Transportation Planning and* Technology, 22 (4), 1999, pp. 271-286

Ficher, M.M., Computational Neural Networks: An Attractive Class of Mathematical Models for Transportation Research, in *Neural Networks in Transport Applications*, V. Himanen, P. Nijkamp, A. Reggiani and J. Raitio (eds.), Ashgate, 1998, pp. 3-20

Florez, H., *Empty-Container Repositioning and Leasing: An Optimization Model*, PhD thesis, Polytechnic Institute of New York, 1986

Franck, M. and Wolfe, P., An Algorithm for Quadratic Programming, *Naval Research Logistics Quaterly* 3, 1956, pp. 95–110

Gambardella, L.M., Mastrolilli, M., Rizzoli, A.E., and Zaffalon, M, An optimization methodology for intermodal terminal management, *Journal of Intelligent Manufacturing* 12(5-6), 2001, pp. 521-534

Gambardella, L.M., Rizzoli E. Funk P, Agent-based Planning and Simulation of Combined Rail/Road Transport, *Simulation* 78(5), 2002, pp. 293-303

Gambardella, L.M., Rizzoli, A.E., and Zaffalon, M, Simulation and Planning of an Intermodal Container Terminal, *Simulation* 71(2), 1998, pp. 107-116

Haghani, A.E., Rail Freight Transportation: A Review of Recent Optimization Models for Train Routing and Empty Car Distribution, *Journal of Advanced Transportation* 21, 1987, pp. 147-172

Harker, P.T., Services and Technology: Reengineering the Railroads, *Interfaces* 25(3), 1995, pp. 72-80

Holguín-Veras, J. and Jara-Díaz, S., Optimal pricing for priority service and space allocation in container ports, *Transportation Research Part B* 33, 1999, pp. 81-106

Holguín-Veras, J., and Walton, C.M., Implementation of priority systems for containers at marine intermodal terminals, *Transportation Research Record* 1602, Transportation Research Board, Washington, DC, 1997a , pp. 57-64

Holguín-Veras, J., and Walton, C.M., *The Performance Analysis of Priority Systems*, Research Report SWUTC/97/467304-1, Center for Transportation Research, Department of Civil Engineering, The University of Texas at Austin, TX, 1997b

Holguín-Veras, J., and Walton, C.M., The role of information technology on the implementation of priority systems, *Proceedings of Third Annual World Congress on Intelligent Transportation Systems*, Orlando, FL, 1996b

Holguín-Veras, J., and Walton, C.M., The state of the practice of information system and information technology at marine container ports, *Transportation Research Record* 1522, 1996a, pp.87-93

Holmberg K., Joborn M., Lundgren J.T., Improved Empty Freight Car Distribution, *Transportation Science* 32(2), 1998, pp. 163-173

Huisken, G. and van Berkum, E, Short-Term Travel Time Prediction Using a Neural Network, in *Proceedings of the 13th Mini EURO Conference*, Polytechnic University of Bari, 2002

Hykin S., *Neural Networks*, Macmillan, 1994, ISBN 0-02-352761-7

Ireland P., Case R., Fallis J., Van Dyke C., Kuehn J., Meketon M., The Canadian Pacific Railway Transforms Operations by Using Models to Develop Its Operating Plans, *Interfaces* 34(1), 2004, pp. 5-14

Jalard, N., Deau, D., and Guigon, M., Les Transbordements Rapides : Système Commuter, *Revue Générale des Chemins de Fer*, April, 1990

Joborn, M., Crainic, T.G., Gendreau, M., Holmberg, K., and Lundgren, J.T., Economies of Scale in Empty Freight Car Distribution in Scheduled Railways, *Transportation Science* 38(2), 2004, pp. 121-134

Joborn, M., *Optimization of Empty Freight Car Distribution in Scheduled Railways*, Ph.D. thesis, Division of Optimization, Department of Mathematics, University of Linköping, Linköping, Sweden, 2001

Jordan, W.C. and Turnquist, M.A., A Stochastic Dynamic Network Model for Railroad Car Distribution, *Transportation Science* 17, 1983, pp. 123–145

Kia, M., Shayan, E., and Ghotb, F., Investigation of port capacity under a new approach by computer simulation, *Computers & Industrial Engineering* 42, 2002, pp. 533-540

Kececi N. and Modarres, M. (1998), Software development life cycle model to ensure software quality, *International Conference on Probabilistic Safety Assessment and Management.*

Kelton, W.D., Sadowski, R.P., and Sturrock, D., (2004) *Simulation with ARENA*, 3rd Edition, McGraw-Hill.

Klodzinski, J. and Al-Deek, H.M., Transferability of an Intermodal Freight Transportation Forecasting Model to Major Florida Seaports, *Transportation Research Record* 1820, 2003, pp. 36-45

Kozan, E., Increasing the operational efficiency of container terminals in Australia, *Journal of the Operational Research Society* 48(2), 1997, 151-161

Kraft, E.R. and Guignard-Spielberg, M., *A Mixed Integer Optimization Model to Improve Freight Car Classification in Railroad Yards*, Report 93-06-06, Department of Operations and Information Management, The Wharton School, University of Pennsylvania, 1993

Kraft, E.R., *A Reservations-Based Railway Network Operations Management System*, Ph.D. dissertation, University of Pennsylvania, Department of Systems Engineering, 1998

Kraft, E.R., The Link Between Demand Variability and Railroad Service Reliability, *Journal of the Transportation Research Forum* 34(2), 1995, pp. 27-43

Kwon, O.K., *Managing Heterogeneous Traffic on Rail Freight Network Incorporating the Logistics Needs of Market Segments*, Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 1994

Lam, W.H.K., Ng P.L.P., Seabrooke W., Hui E.C.M., Forecasts and reliability analysis of port cargo throughput in Hong Kong, *Journal of Urban Planning and Development*, ASCE 130(3), 2004, pp. 133-144

Lee, D.H., Song, L., and Raguraman, K., *An Optimization Approach for Seaport Security Operations*, Technical report, Transportation and Logistics Security Group, The Logistics Institute, Georgia Tech, 2003

Lewis, B.M., Erera, A.L., and White III, C.C., Efficient Container Security Operations at Transshipment Seaports, in *Proceedings of the Inaugural Annual Intelligent Transportation Society (Singapore) Symposium*, September 5-6, 2002, Singapore

Lewis, B.M., Erera, A.L., and White III, C.C., Optimization Approaches for Efficient Container Security Operations at Transshipment Seaports, *Transportation Research Record* 1822, 2003, pp. 1-8

Li, S.-T., and Shue, L.-Y., A study of logistics infomediary in air cargo tracking, *Industrial Management & Data Systems* 103/1, 2003, pp. 5-13

Lyons, G.D., Hunt, J.G., and Yousif, S.Y., Factors Influencing the Performance of a Neural Network Driver Decision Model: A Case Study Using Simulated Data, *Neural Networks in Transport Applications*, edited by V. Himanen, P. Nijkamp, A. Reggiani and J. Raitio, Ashgate, 1998, pp. 229-248

Macharis, C. and Bontekoning, Y.M., Opportunities for OR in Intermodal Freight Transport Research: A Review, *European Journal of Operational Research* 153, 2004, pp. 400-416

Malavasi, G. and Ricci, S., Simulation of stochastic elements in railway systems using self-learning processes, *European Journal of Operational Research* 131(2), 2001, pp. 262-272

Marlow, P.B., and Paixão Casaca, A.C., Measuring lean ports performance, *International Journal of Transport Management*, vol. 1, 2003, pp. 189-202

Martinez, F.M., Gutierrez, I.G., Oliveira, A.O., and Bedia, L.M.A., Gantry crane operations to transfer containers between trains: A simulation study of a Spanish terminal, *Transportation Planning and Technology* 7(4), 2004, pp. 261-284

Martland, C.D., Little, P., and Sussman, J.M., Service Management in the Railroad Industry, in *Railroad Freight Transportation Research Needs*, Conference Proceedings 2, of the joint AAR/FRA/TRB conference in Bethesda, Md, July 12-14, 1993, National Academy Press, Washington D.C., 1994, pp. 89-103

Mccalla, R. *Water transportation in Canada*, Halifax: Formac, 1994, 259 p.

Mercer Management Consulting, The Service Quality Challenge for the 1990s, *The 5th American Railroad Conference*, November 1991

Merkle, E., *The Criticality of Meteorological Data on Port Operations as Ships are Getting Deeper and Taller*, conference paper, American Meteorological Society, New York, 2004

Missouri Pacific Railroad, *Missouri Pacific's Computerized Freight Car Scheduling System: Functional Requirements*, Report number FRA/OPPD-77/10 Final Rpt, NTIS Order Number PB-2754239/8ST DOTL NITS, 1977

Morlok, E.K. and Chang, D.J., Measuring capacity flexibility of a transportation system, *Transportation Research Part A* 38, 2004, pp. 405-420

Morlok, E.K., *Concepts and Measurements of Transportation System Flexibility*, Working paper, Electrical and Systems Engineering Department, University of Pennsylvania, Philadelphia, PA, 2003

Olivier, Daniel *Problèmes de capacité et stratégies d'expansion au port de Montréal*, Mémoire de M.Sc. en géographie, Université de Montréal, 2002, 35 p.

Park, D.J. and Rilett, L.R., Forecasting multiple-period freeway link travel times using modular neural networks, in *Land Use and Transportation Planning and Programming Applications*, *Transportation Research Record* 1617, 1988, pp. 163-170

Park, Y.M. and Kim, K.H., A Scheduling Method for Berth and Quay Cranes, *OR Spectrum* 25, 2003, pp. 1-23

Potvin, J.-Y. and Smith, K.A., Artificial Neural Networks for Combinatorial Optimization, rapport CRT-2001-02, Centre de recherche sur les transports, 2001

Powell, W. B. and H. Topaloglu, Stochastic Programming in Transportation and Logistics, in *Handbooks in Operations Research and Management Science: Stochastic Programming*, Shapiro, A. and Ruszczynski, A. (eds.), Elsevier, Amsterdam, 2003, pp. 555-635

Powell, W.B. and Carvalho, T.A., Real-Time Optimization of Containers and Flatcars for Intermodal Operations, *Transportation Science* 32(2), 1998, pp. 110–126

Powell, W.B. and H. Topaloglu, Fleet Management, in *Applications of Stochastic Programming*, Wallace, S. and Ziemba, W. (eds.), Math Programming Society – SIAM Series in Optimization, 2005, 185-216.

Powell, W.B., Bouzaïene-Ayari, B., and Simaõ, H.P., Dynamic Models for Freight Transportation, in *Handbooks in Operations Research and Management Science: Transportation*, Barnhart, C. and Laporte, G. (eds.), Elsevier, Amsterdam, forthcoming, 2005

Powell, W.B., Dynamic Models of Transportation Operations, in *Handbooks in Operations Research and Management Science: Supply Chain Management*, Graves, S. and de Tok, T.A.G., (eds.), Elsevier, Amsterdam, 2003, pp. 677-756

Prognos, A.G., *Marché Européen des Transports de Marchandises*, Bâle, December 1998

Psaraftis, H.N., Foreword to the focused issue on maritime transportation, *Transportation Science* 33(3), 1999, pp.1-2

Riedmiller, M., and H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, *Proceedings of the IEEE International Conference on Neural Networks*, 1993.

Ronen, D., Cargo ships routing and scheduling: Survey of models and problems, *European Journal of Operational Research* 12, 1983, pp. 119-126

Ronen, D., Ship scheduling: The last decade, *European Journal of Operational Research* 71(3), 1993, pp. 325-333

Sadek A.W., Spring, G., and Smith B.L., Toward more effective transportation applications of computational intelligence paradigms, *Initiatives in Information Technology and Geospatial Science for Transportation, Transportation Research Record*,1836, 2003, pp. 57-63

Sarvareddy, P., Al-Deek, H.M., Klodzinski, J., and Anagnostopoulos, G., Evaluation of Two Neural Network Methods for Generating Heavy Truck Trips at Port Canaveral using Vessel Freight Data, submitted to *Transportation Research Record*, 2004

Sayed T, and Razavi, A., Comparison of neural and conventional approaches to mode choice analysis, *Journal of Computing in Civil Engineering* 14(1), 2000, pp. 23-30

Schinas, O., Lyridis, D.V., and Psaraftis, H.N., Introducing E-brokerage in European Transport Services: the Case of the PROSIT Project, in *Proceedings of The Fifth International Conference on Electronic Commerce Research* (ICECR-5), Montreal, Canada, October 23-27, 2002

Sgouridis, S.P. and Angelides, D.C., Simulation-based analysis of handling inbound containers in a terminal, in *Proceedings of the 2002 Winter Simulation Conference*, Yücesan, E., Chen, C.-H., Snowdon, J.L., and Charnes, J.M. (eds.) 2002, pp. 1716-1724

Shaeffer, W.W., and Stern, G.L., The Influence of Locomotive Reliability upon Locomotive Acquisition and Maintenance Policies, *Proceedings of the 27th Annual Meeting, Transportation Research Forum*, 1, 1986, pp. 113-114

Shmueli, D, Salomon I., and Shefer, D., Neural Network Analysis of Travel Behaviour, *Neural Networks in Transport Applications*, edited by V. Himanen, P. Nijkamp, A. Reggiani and J. Raitio, Ashgate, 1998, pp. 151-169

Steenken, D., Voß, S., and Stahlbock, R., Container terminal operation and operations research – a classification and literature review, *OR Spectrum* 26, 2004, pp. 3-49

TR 1033 INTERPORT, Telematics Application Programme, *Integrating Waterborne Transport in the Logistics Chain*, Final Report, 2000, 31p.

Transport Canada, *Impact de l'introduction des SVCEI sur la fluidité du transport maritime sur le Saint-Laurent*, TP 13092F, Ottawa, 2003

Transportation Research Board, 2002, www.customs.ustreas.gov/xp/cgov/home.html, U.S. Customs and Border Protection, Department of Homeland Security

UNCTAD, *Container Security: major initiatives and related international measures*, Geneva, 2004, 42 p.

Vis, I.F.A. and de Koster, R., Transshipment of containers at a container terminal: An overview, *European Journal of Operational Research* 147, 2003, 1-16

White, W.W. and Bomberault, A.M., A Network Algorithm for Empty Freight Car Allocation, *IBM Systems Journal* 8(2), 1969, pp.147–171

White, W.W., *A Program for Empty Freight Car Allocation*, Technical Report 360D.29.002, IBM Contributed Program Library, IBM Corporation, Program Information, Department, Hawthorne, N.Y., 1968

White, W.W., Dynamic Transshipment Networks: An Algorithm and its Application to the Distribution of Empty Containers, *Networks* 2(3), 1972, 211–236

Yusuf, Y.Y., Sarhadi, M., and Gunasekaran, A., Agile manufacturing: the drivers, concepts and attributes, *International Journal of Production Economics* 62(1–2), 1999, 33–43

Annex B – Technical Specification of Simulation Model

This Annex presents the technical document describing the implementation of the CN network using the first-generation simulator.

CN simulated network model: technical description

This document describes the implantation of the part of the CN network for which we simulate the operations. This documentation is divided into seven sections. The first one is a brief description of the physical network on which our simulation model is built. In this section we present the terminal, the rails and the services that we are simulating along those rails and terminals. The following four sections describe the main model by explaining each one of the four submodels it is made of. Following the four sections describing each submodel, we have a section describing the simulated network initialization. Before the last section, we have included a section presenting the optimization aspect that we have developed to enhance our Arena model. Finally we have a section making the list of assumptions that have been made while developing this model. Although this document is not an introduction to Arena, we have put throughout the document sections called "Arena technical vocabulary", introducing Arena concepts when needed.

1. The CN simulated network: physical description

The simulated network consists of a section of the North American CN network, limited to six terminals and the rails connecting those terminals.

1.1. The terminals

The six terminals that are considered are: CHI (located in the city of Chicago), BIT (located in the city of Toronto), MTL and POM (located in the city of Montreal), MCT (located in the city of Moncton) and Halifax (located in the city of Halifax). Note that into the Halifax terminal we merge three terminals, HIT, HHL and FCV. We refer to those six terminals as the "simulated terminals". The reason behind this appellation is that we simulate train activities in them, such as train arrival, train processing, train make up and train departure.

As opposed to the simulated terminals we consider "non simulated terminals", in which we don't simulate train activities because the train related to our four services don't pass through those terminals. We have three non simulated terminals: DET, Auburn and one that we call Virtual MTL. In those terminals, we only simulate loading and blocking activities (no train arrival, processing or departure). The blocks resulting from those operations are directly sent to the terminal where they appear in the block plan. In reality, those blocks are brought to the terminal where they appear in the block plan by a service that we don't simulate. The third virtual terminal is called Virtual MTL because it receives a part of the demand of MTL that is being processed for non simulated services that carry this demand to BIT, where it is confide to our simulated services. Since we don't simulate the service that carries this demand from MTL to BIT, we generate and process the demand in Virtual MTL, then send it directly to BIT, like the demand originating from Auburn and DET.

1.2. The rails

There are rail connections, in both ways, between CHI and BIT, between BIT and MTL, between MTL and POM, between MTL and MCT and finally between MCT and Halifax. Those rail connections imply that a train can only reach POM by reaching first MTL.

1.3. The services

We take into consideration four services: service 120, service 121, service 148 and service 149. Origins and destinations are respectively BIT and Halifax for service 120, Halifax and BIT for service 121, CHI and Halifax for service 148, Halifax and CHI for service 149. The simulated part of the CN network is divided into a Western section, corresponding to the terminals CHI, BIT, MTL and POM, and into an Eastern section, corresponding to Halifax, MCT, MTL and POM. MTL and POM make a junction between the Eastern part and the Western part, so they appear in both sections.

The engines related to services 148 and 149 are affected to one of those two sections as follows: an engine originating from CHI and related to service 148 passes through BIT, and MTL, labelled as service 148 and then returns to CHI, passing through POM and BIT, labelled as service 149; an engine originating from Halifax related to service 149 passes through MCT and POM, labelled as service 149, and then returns to Halifax, passing through MTL and MCT, labelled as service 148.

We don't have to take into consideration the Eastern/Western particularity of the CN network for services 120 and 121: an engine originating from BIT with service 120 goes through POM, MCT and Halifax, labelled as service 120; and an engine originating from Halifax with service 121 goes through MCT, POM and BIT, labelled as service 121.

1.4. Implantation of the simulated network

The model has been developed according to the physical description made above and has been developed with Arena 8.01 under Windows XP.

The main model is divided into four submodels: Demand submodel, Simulated Network submodel, Delivery submodel and Technical Issues submodel. Each submodel is described below.

2. Technical Issues submodel

In this submodel we take care of technical details underlying the simulation of the CN network. This submodel is divided into six submodels: in three of them, "Attributes", "Entities" and "Variables", we declare and initialize all types of variables called in our model. The fourth one, "Initializations", initialize all data structures in our model. In the last ones, "Parameter Updates for Simulated/Non Simulated Terminal", we update parameters underlying the different actions in the terminals (loading, blocking, train departure). But before describing those submodels, we will define some technical Arena vocabulary that we will use throughout this document.

2.1. Arena technical vocabulary: entities, modules, connections

In any Arena model, we must create what are called "entities". Without creation of an entity, a simulation is not possible. Entities modify the event calendar on which the simulation is based. To modify the event calendar, entities enter what are called "modules" that describe an action (for example, a "delay" action or a "seize resource" action). Creating an entity is also made with a module: a "Create" module. To move from one module to another, entities use "connections" that are represented with lines. They are the paths, leading entities through a succession of modules. The last module any entity must enter is a "Dispose" module where it is destroyed.

2.2. Entities submodel

When an entity is created, we must fill a field called "Entity Type". An "Entity Type" is attached to each entity to identify its nature. In the Entities submodel we define every possible entity type in our model. We consider five types of entities in our model: "Container Type Entity", "Car Type Entity", "Engine Type Entity", "Block Type Entity" and "Train Type Entity" (we will see in the "Simulated Network submodel" section how we can obtain a "Block Type" entity or a "Train Type" entity from the other types of entities). Inspection of the "Entity Type" is sometimes needed to decide what actions are to be taken. For example, when loaded cars arrive to their destination and are unloaded, we have to deal with containers and cars. If "Entity Type" is "Container Type Entity", we must deliver a container but if "Entity Type" is "Car Type", we must send a car back to the stock of empty cars of the current terminal. We will explain latter how we obtain a loaded car from a car and container(s) and vice versa.

2.3. Data structures: Attributes and Variables

The data structures are defined into two submodels, depending on their nature: attribute or variable.

2.3.1. Arena technical vocabulary: variables and attributes

"Attributes are entity dependent. Each entity has the same set of attributes, but the value of any attribute does not have to be the same among different entities. Variables contain global values that look the same to all entities at any point in time." (Arena Online Documentation).

By default, Arena initializes attributes and variables to zero, but our "Attribute" and "Variables" submodels allow us to initialize attributes and variables to other values. Also, those two submodels allow us to define arrays of attributes and variables (one or two dimension arrays)

2.3.2. Attributes submodel

In the "Attributes" submodel we define all the entities attributes that are used in the model. Once defined in this submodel, we access a particular attribute with a "drop down" list, for example when we need to check its value or assign a particular value to it. Examples of attributes are:

- "Car Origin",
- "Car Destination",
- "Car is Loaded" (evaluates to variables "True" (=1) or "False" (=0)),
- "Block Tag",
- "Block is Already on Train" (True or False),
- "Block Current Length in Feet", ...

Note that in an Arena model the list of attributes is the same for all the entities. So, in our model, when we consider a "Car Type" entity for example, we are interested only in the attributes concerning cars.

2.3.3. Variables submodel

In the "Variables" submodel, we define all the variables that we need to develop our model. We classify them according to what they are related to, into thirteen categories:

- General. This category includes basic variables not directly related to any other category:
  - the boolean "True" = 1 and "False" = 0,
  - an epsilon value "EPS" = 0,0001,
  - the number of hours defining a day "Hours Per Day" = 24,
  - the moment the simulation starts "Simulation Starts" = 0,0005,
  - the number of days to simulate "Nb Days to Simulate".
- Reading Files. This category includes all the variables used to read files when initializing data structures. For example, temporary variables used to read a terminal or a service in a data file, counters, ...
- Network. This category contains variables describing the network, such as the travel times between terminals.
- Destinations. We might as well have named this category "Origins": its contains all the variables describing the terminals, such as the number of destinations, the destinations (West, CHI, DET, BIT, MTL, POM, Auburn, MCT, HIT, HHL, FCV), the number of simulated terminals, the simulated terminals, (CHI, BIT, MTL, POM, MCT, Halifax), the terminal corresponding to each destination (this terminal is Halifax for the three destination HIT, HHL and FCV and is the same as the destination for all other destinations).
- Demand Generation. This category includes all variables used when generating containers.
- Services. This category includes the variables describing the simulated services: the number of services and the four services (Service 120, Service 121, Service 148 and Service 149).
- Containers. This category contains variables used to describe containers. For example, the number of containers types, the container types ("Twenty Feet Empty", "Twenty Feet Light", Twenty Feet Medium", "Twenty Feet Heavy", "Forty Feet Empty", "Forty Feet Light", "Forty Feet Medium", "Forty Feet Heavy"), the thresholds used to determine a container type from its weight.
- Cars. This category contains variables used to describe the cars: number of car types, car types, number of cars in stock, car length, ...
- Trains and Engines. This category includes variables concerning trains and engines, such as the number of available engines in a terminal, the train length in blocks when leaving a terminal, ..

- Blocks. This category includes variables describing blocks: for example, to define the blocks, we have the array of blocks "Blocks" holding our blocks tag (1 to 34). Then, we have an array of origins ("Block Origin") associated with each of our 34 blocks and an array of destinations ("Block Destinations"). The array "Block Length in Feet" holds the capacity of each block in feet. We also have other variables regarding blocks to describe the services that take care of moving them from their origin to their destination.
- Waybills. Waybills are also defined by several variables: "Nb Waybills", the number of waybills, "Waybill Lengths", the number of blocks defining each way bill, "Waybill Blocks", the blocks defining each waybill. Since many cars share the same waybill, we have defined them as variables (global) to be able to access them from any entity. One of the reasons for keeping the waybills global is memory efficiency.

So, with those data structures, how do we keep track of a car on its waybill? Each car has an attribute called "Container or Car Waybill" (the way bill is the same for a container and the car that carries it), which is a tag between 1 and 49 (there are 49 possible waybills). Global data structure information corresponds to this tag, such as "Waybills Length(Container or Car Way Bill)", ... To determine on which block of the waybill is the car, we need to check the car attribute "Car Current Transfer": for example, if "Car Current Transfer" is zero, the car is on the first block of the way bill. Typically, when we "dislocate" a block in the simulation to check whether the cars are at their destination or not, when a car has not reached its destination, we increment its attribute "Car Current Transfer" by 1 and assign the car attribute "Car Current Block" to the block of the waybill corresponding to that transfer, which is the next in the array Waybill Blocks corresponding to the current waybill. So, to summarize, the waybills are global information, so accessible by any entity, and each entity has information about the waybill tag and the transfer number, allowing it to know at any time where it is on its waybill.

- Loading and Blocking. This category includes variables underlying loading, blocking and train departure activities in each terminal
- Signals (Arena technical vocabulary: "signals").

In Arena, it is possible to interrupt the course of an entity in a "Hold" module. When entering this module, an entity then waits for a condition to be released from the module. This condition can be an expression that must evaluate to true, or the reception of a "signal". Upon reception of this signal, a variable indicates how many entities (if more than one in the "Hold" module) car be released

The signal category contains all the signals used in our model, such as signals used to release containers for loading, to release empty cars for loading and to release empty cars for blocking.

2.4. Initializations

In the Initialization submodel, we initialize all data structures used in the model. It is divided into eight submodels corresponding to the different data structures.

2.4.1. Blocks Initialization

We initialize the blocks from the file Blocks CN.txt. In this file, we have, for each block:

- the block number (a tag between 1 and Nb Blocks = 34)
- the block origin
- the block destination
- the block length (in feet)
- the block number of service transfers
- if the block number of service transfers is greater than zero
    - for each service transfer:
        - terminal where the block is transferred
        - service to which the block is transferred
    - for each simulated terminal
        - terminal
        - service carrying the block when leaving the terminal

Although the information concerning block service seems redundant (we could have deduce the service transfers from the list of terminals/service carrying the block in the current terminal) it is not. For example, a block coming from Halifax with Service 149 going to CHI is considered to be transferred in POM to Service 149 (the same service) since engines of service 149 stay in the Eastern section of the network and leave their blocks to an engine of the Western section of the network.

When initializing the blocks, we determine in which terminal a block is formed, by inspecting the block origin. So, when we are done initializing the blocks, we have for each terminal a list of blocks that are formed in the terminal.

2.4.2. Waybills Initialization

A Waybill is defined by a succession of blocks. Waybills are read in the file Waybills CN.txt where, for each waybill we have:

- the waybill number (a tag between 1 and Nb Waybills = 49)
- the waybill length (it is the number of blocks defining the waybill)
- the blocks defining the waybill

Waybills in this file have been built by considering all the possible successions of blocks in the network.

As waybill are read, we build a two dimensional array (OD to Waybill) that has as the first index an origin, as the second index a destination and as the corresponding value to this pair of index a waybill. We use this array to associate a waybill to a container entering the system from its origin and its destination. If more than one waybill can carry a container between its origin and its destination, only the shortest one is considered (see note 6).

2.4.3. Combinations Containers Cars Initialization

The possible combinations between containers and cars are needed when it is time to load containers on empty cars. At this moment, we need to choose among the different available car types and the different types of containers the ones that we are going to match to build a loaded car. With very little information from the CN on this subject, we built a list of possible combinations, from the preferred one to the one that we will avoid if possible with the idea to match first the containers that are the most difficult to load. We have 59 possible combinations that are read in the file Combinations Cars Containers CN.txt. Each possible combination is described by the number of each component it contains:

- number of "Twenty Feet Empty" containers
- number of "Twenty Feet Light" containers
- number of "Twenty Feet Medium" containers
- number of "Twenty Feet Heavy" containers
- number of "Forty Feet Empty" containers
- number of "Forty Feet Light" containers
- number of "Forty Feet Medium" containers
- number of "Forty Feet Heavy" containers
- number of "Low Capacity Single Stack" cars
- number of "Low Capacity Double Stack" cars
- number of "High Capacity" cars

Low capacity cars can carry up to 118000 pounds and high capacity cars up to 160000 pounds.

While considering those possible combinations, we keep in mind that we have a target proportion of high capacity cars to use. So if we consider a combination using a high capacity car when our target proportion is reached, we skip to the next possible combination.

2.4.4. Loading Blocking Flags Initialization

For each terminal, we have flags indicating the possible activities in the terminal, for each simulated service. They are read in the file Loading Blocking in Terminals CN.txt, where we have, for each simulated and non simulated terminal:

- the terminal
- for each service:
  loading in terminal for concerned service (True or False)
  blocking in terminal for concerned service (True or False)
  train departure from terminal for concerned service (True or False)

For example, there are no loading or blocking activities in MCT for Service 14, but there is train departure from MCT for this service.

2.4.5. Signal Initialization

The only initialization performed without a file is the signal initialization. We simply initialize every signal used in the model to a unique value from 1 to the number of signals.

2.4.6. Terminal Destination Initialization

As we have seen earlier, there is a difference between a destination and a terminal when we merge several destinations into one. It is the case only for Halifax, where this terminal represents three destinations: HIT, HHL and FCV. The terminal destinations are initialized from the file Terminal Destinations CN.txt where we have, for each destination, the corresponding terminal. Note that for the destination West, we do as if there was a terminal called West

2.4.7. Train Time Table Initialization

Train time tables are initialized from the file Train Time Table CN.txt where there is, for each terminal:

- the terminal
- departure time of service 120
- departure time of service 121
- departure time of service 148
- departure time of service 149

When there is no departure from a terminal of a service (for example there is no departure of service 120 from CHI, since this service originates from BIT), the value entered for this field will be ignored since the corresponding flag (Departure in Terminal) will be set to False. All the train departures are set to the Eastern time (time of BIT,MTL and POM).

All train departure times were provided by the CN, except for the ones from MTL to POM and vice versa for services 148 and 149. For those two services, the engines are associated to the western part of the network (CHI, BIT, MTL, POM and back to BIT and CHI) or to the eastern part of the network (Halifax, MCT, POM, MTL and back to MCT and Halifax). For the engines associated to those services, we have set a delay of 2 hours for train operations in MTL and POM (based on other delays allowed for train operations in other terminals) and a travel time of 20 minutes between MTL and POM.

2.4.8. Travel Times Initialization

Train Travel times have been deduced from the train departure and arrival times in each terminal. They are read in the file Travel Times Between Terminals CN.txt where we have, for each simulated terminal:

- the terminal
- the number of adjacent terminals
- for each adjacent terminal
    - the adjacent terminal
    - the number of services making the connection from the original terminal to the adjacent one
    - for each service making the connection:
        - the service
        - the travel time

2.5. Updating Parameters for (Non) Simulated Terminals

Let remind us that by "simulated terminals" we mean one of the six terminals defining the simulated network: CHI, BIT, MTL, POM, MCT and Halifax. By "Non Simulated Terminal" we mean a terminal that is not a part of the simulated network, but that we must take into consideration because we deliver through our simulated network a part of the demand that originates in those terminals. We have three "Non Simulated Terminals": Auburn, DET and one that we call "Virtual MTL" (from which a non simulated service delivers blocks to BIT). The "updating Parameters for (Non) Simulated Terminals" submodels are divided into three submodels: Loading Activities, Blocking Activities and Departure Activities.

2.5.1. Loading Activities

In this submodel we take into consideration the containers to be loaded and the available empty cars. When it is time to load cars (a fixed time before train departure) signals are sent to release containers and cars, to form loaded cars. The combinations of cars and containers are determined according to a set of possible combinations, using a target proportion of high capacity cars that we want to use while loading. This porportion is the percentage of high capacity cars that we want to have in the block that will be formed with the loaded cars. A loaded car is sent to the next activity, blocking.

2.5.2. Blocking Activities

Here we take into consideration the loaded cars, the available empty cars and the block lengths. When it is time to block cars, loaded cars forming a same block are grouped and signals are sent to the stock of empty cars to complete it (if needed), still using the target proportion of high capacity cars that we want to be part of the block.

2.5.3. Departure Activities

In this submodel we update variables used in terminals to determine train departure.

3. Demand submodel

The demand submodel is divided into two submodels: container generation and block generation.

3.1. Container generation

The first one "Generate Containers" generate the demand in form of containers for every terminal, simulated or not. This generation of containers is performed by reading containers description from a file. We assign values to each of the following container attributes:

- Container Origin
- Container Destination
- Container Length
- Container Weight
- Container Type
- Container or Car Waybill

We use thresholds to classify the containers in four categories depending on their weight. Eight container types are possible and are described below (see note (5)).

Container Type	Merchandise weight in pounds
Twenty Feet Empty	0
Twenty Feet Light	0 - 35000
Twenty Feet Medium	35000 - 50000
Twenty Feet Heavy	50000 - 56000
Forty Feet Empty	0
Forty Feet Light	0 - 40000
Forty Feet Medium	40000 - 60000
Forty Feet Heavy	60000 - 65000

The file from which the containers are generated is in the following format:

- Number of days to read in the data file
- Number of Origin/Destination combinations to read in the data file

For each day

- Current day

For each OD combination

- Origin of current OD combination
- Destination of current OD combination
- Number of containers for current day and current OD combination

For each container of current day and current OD combination

- Container Length (20 or 40 feet)
- Container Weight (in pounds)

3.2. Block generation

Although the demand is generally expressed in form of container arrival in a terminal, it is also expressed as block arrival for BIT terminal. In this terminal, we have three blocks coming from other terminals with a non simulated service: POMxDET (500'), CHIxMTL (1800') and CHIxAuburn (200'). In the block generation submodel, we defined three "non simulated" terminals (DET, Auburn and Virtual MTL) with only loading and blocking activities, performed using containers generated for those three terminals in the Container Generation submodel described in the previous section. Since we don't simulate the service carrying those blocks from their origin to BIT, once the blocks are formed, they are directly sent to the demand arrival part of BIT.

4. Simulated Network submodel

4.1. General description of the Simulated Network

Arena technical vocabulary: "routes"

In Arena, we can connect modules with "routes". A "route" is a connection to which we can associate a travel time.

The "physical" CN network is represented by five cities (five submodels: Chicago, Toronto, Montreal, Moncton, Halifax) connected by Arena "routes" representing the CN lanes. If animation is turned on, we should see trains moving along the routes, to which we have associated the train travel time, from one city to another. Cities of Montreal and Halifax are divided into many terminals: in Montreal, we have MTL and POM, and in Halifax we have HIT, HHL and FCV. Every other city contains only one terminal. The terminal submodel is developed in a "generic" way, so that the same model can be adapted for each specific case. For example, when a train is entering a terminal, the train attribute "Current Terminal" is set to the name of this terminal. After that, whenever we need to access the name of the terminal within this terminal, we do so by inspecting the Arena attribute "Current Terminal" of the Arena entity we are taking into consideration. Each terminal is divided into many submodels defining specific operations on containers, blocks and trains. We will describe these submodels in a following section.

4.2. Entities moving in the network

The smallest entities that we have in our simulated network submodel are the containers to be delivered, the empty cars and the engines. We obtain another entity, the loaded car by merging containers with an empty car when simulating the loading process. Once a loaded car is created, we can split it into its original entities to get back the empty car and the containers to be delivered. Following this logic, we obtain a block by merging several cars (loaded or empty) and a train by merging an engine and several blocks. So, when we say that containers, empty cars and engines are the "smallest entities", we mean entities that are not obtained by merging other entities.

4.3. A generic terminal

A terminal has two "entry points", one for train arrival and one for demand arrival; a terminal has also two "exit points", one for train departure and one for container delivery. Different submodels make the connection between entry and exit points. We describe the nine submodels defining all the operations that take place in a terminal below.

4.3.1. Train Arrival submodel

Here is where a train entity enters the terminal. We assign its attribute "Current Terminal" to the name of the current terminal and hold the train until the previous train of the same service that has entered the current terminal has been inspected. By "inspected" we mean that the blocks (on this previous train) that needed to be disconnected (because they were at their destination or because they needed to be transferred) have been disconnected from their train.

4.3.2. Train Processing submodel

After leaving this first module of the terminal, the train enters the Train Processing submodel to determine what operations are to be applied. Every block that has reached its destination or that has to be transferred is disconnected from the train. If a block is not yet at its destination nor needs to be transferred, it remains attached to its engine (in fact, in Arena, the train has been totally split (into its engine and its blocks) after entering the Train Processing module and the blocks that needs to remain attached to the engine are then merged together with the engine). Since this operation (splitting the train into its engine and blocks) is purely virtual, no time is associated with it. The engine and attached blocks (if any) are held until the selected blocks are disconnected and are then sent to the "Train Makeup, Passing Train, Out of Zone Train" submodel : it becomes available to form a (new) train. Note that in the case of a passing train, no block disconnection is performed and the whole train is sent to the next submodel. If a block has to be transferred, we assign its attribute "Current Service" to its new service and send the block to the "Train Makeup, Passing Train, Out of Zone Train" submodel. If a block has reached its destination, we split it into the cars it is made of and see if the cars are also at destination. If this is the case, we send them to the Terminal Delivery submodel, and if not we send them to the Classification Blocking submodel.

4.3.3. **Terminal Delivery submodel** (Not yet implemented)

In the Yard Delivery submodel we essentially split (if loaded) cars into an empty car and the containers that are delivered and collect statistics on deliveries before sending the empty car back to the Terminal Stock of Empty Cars submodel and the containers to the "main" Delivery submodel where the entities are disposed.

4.3.4. Classification Blocking submodel

We have seen above that the Classification Blocking submodel is fed with loaded cars disconnected from a train in the Train Processing submodel, but it is also fed by loaded cars corresponding to the localdemand local to the terminal (arriving by trucks for example) and by empty cars coming from the Terminal Stock of Empty Cars submodel. Loaded cars arriving to Classification Blocking submodel are delayed until it is time for blocking cars in the current terminal: a loaded car has a waybil associated to it so we know at any time in the simulation which service the car is attached to with an attribute called "Current Service", and we have a (global) variable Blocking Time (Current Terminal, Current Service) indicating the next blocking time for the current yard and the current service. (Note that this variable needs to be updated once the time Blocking Time(Current Terminal, Current Service) is reached; this is done in the Updating Parameters for Simulated Terminals submodel). Once it is blocking time, the loaded cars are classified according to the current block they are associated to and blocks are filled with them. Note that if there are more loaded cars than the block can receive, the remaining loaded cars will have to wait until the next blocking time to have a chance to be blocked. Once we have tried to fill a block to its capacity with loaded cars, if the capacity is not reached, we try to complete the block with empty cars. Then we merge the loaded (and empty) cars to make a block. The block is then sent to "Train Makeup, Passing Train, Out of Zone Train" submodel.

4.3.5. Train Makeup, Passing Train, Out of Zone Train submodel

Two kind of entities enter this submodel: the engines (eventually attached to some blocks that were on the train when it arrived to the current yard and that were not at destination nor transferred) and the blocks that need to be attached to a train to leave the current terminal.

Engines are held in according to the service they correspond to until the previous train of the corresponding service is made up.

Blocks entering this submodel are held in according to their "Block Tag" attribute until an engine corresponding to the service they are attached to becomes available. We then form the new train one block at a time. Once we have all the blocks and an engine to form a train we merge all the entities into a train entity and send it to the Train Departure submodel.

4.3.6. Train Departure submodel

Here we hold trains until conditions for train departure are reached.

4.3.7. Demand Arrival submodel

As we said above, the Classification Blocking submodel is fed by local demand. This demand arrives in this submodel and is sent to the Demand Processing submodel. Containers arriving to this submodel have a waybill attached to them. In the case of BIT, demand also arrives as blocks, coming from our three non simulated terminals.

4.3.8. Loading submodel

Here we perform the loading operation of containers on empty cars. So this submodel is fed by containers arriving from the Demand Arrival submodel and by empty cars coming from the Terminal Stock of Empty Cars submodel. Containers are classified according to their waybill and to their Arena Attribute "Container Type". To load the containers of a specific waybill, we determine (in the Updating Parameters for (Non) Simulated Terminals submodel), by inspecting the types of containers to be loaded and the types of empty cars available, which containers to match with which car. We then release a car from the Stock of Empty Cars submodel to be merged with the selected containers. The different possible combinations of containers and cars are determined in advance and tried in a systematic way. We use a target value that represents the percentage of high capacity cars that we want when loading a block (that is the first block of the current container's waybill). If the current percentage of high capacity cars is above this target value and if the current combination requires a high capacity car, the combination is refused and we try the next one.

4.3.9. Terminal Stock of Empty Cars submodel

Here we keep all the empty cars of the terminal and release some of them when it is time to load containers or time to block (completion of a block with empty cars). We need to know how the CN completes blocks with empty cars. We consider two types of cars: "Low Capacity Cars" and "High Capacity Cars", which can handle, respectively 118000 pounds and 160000 pounds (see note (2)). Cars are classified according to those types, waiting for a signal to be released. Upon reception of this signal, a flag indicates whether it is time to load containers or time to block cars. In the Updating Parameters for (Non) Simulated Terminals submodel we update variables describing the demand for empty cars for loading operations or blocking operations.

When it is time to load cars, the empty cars are released one at a time by type depending on what needs to be loaded. After being released, we assign a way bill to each car and send them to the Demand Processing submodel where we load containers.

If empty cars are needed for blocking, we have variables indicating how many cars are needed for each block formed in the terminal, so we can assign a block to each car and send them to the Classification Blocking submodel. The destination assigned to each empty car sent to blocking is the corresponding block destination (see note (1)).

5. Delivery submodel

In this submodel that collect statistics regarding the time each container has spent in the system. More precisely, for every container that is being delivered, we write a new line into an output file on which we have the container origin, the container destination and the number of days spent since the container arrival day in its origin terminal. This output file is then processed after the simulation has stopped.

6. Simulated network initialization

The first difficulty in the initialization of the network is that we cannot simply place engines in each simulated terminal, because in some terminal, trains leave with blocks that they had before arriving to this terminal. Since a train must be complete (in terms of the number of blocks it is made of) before leaving a terminal, we would have incomplete trains that couldn't leave a terminal if we initialized the network as we just said. So, at initialization, we can only place engines in terminal where a completely new train is built.

This imply that engines related to services 120 and 121 can only be placed initially at those services origins. For services 148 and 149, although it seems, following this logic, that it is only possible to place engines initially at extreme terminals (CHI, POM and Halifax), it is also possible to place engines in MTL where the engine of the train labelled service 149 is detached to return to Halifax, labelled service 148. For the section of service between MTL and POM it is useless to initially put an engine in POM because the train formed in POM must wait in MTL for a block coming from Halifax, brought to MTL by service 149. According to the train time tables, this block will arrive at MTL after the engine coming from CHI related to service 148 will have reached MTL, POM and then MTL again.

Once we have placed engines to their initial terminal, we face a second difficulty. It is obvious that we cannot generate demand for a terminal before an engine (corresponding to the service that will carry that demand) has initially reached this terminal. The reason for this is that this demand would "wait" for an initial engine to reach the concerned terminal, and so its delivery time would be increased by this waiting time that does not exist in reality. Delaying the demand (more precisely, ignoring the demand for a certain number of days, until every terminal has been reached by at least one engine) is part of the solution, but not the complete solution. Why? Because, even if we do not have demand generated for a terminal, there are still blocking operations taking place periodically in this terminal. So, we would accumulate blocks made of empty cars until an engine reaches the terminal, and the demand arriving would be delayed by the time it takes to evacuate all those blocks. Once again, we introduce a delay that does not exist in reality.

So the solution is to inhibit (loading and) blocking operations in the terminals until the arrival of the first engine. More precisely, during the simulation, we inspect the first train departure time for each terminal and each service, and, when it is time for (loading or) blocking operations, we check if the current operation is related to (or greater than) this train departure time. If it is the case, the current operation takes place, otherwise it is skipped. Then, we only generate demand when the network is operational, which means when all terminals have been reached by at least one engine of each service it is supposed to be reached by.

7. Block Length Optimization

Although we can simulate the CN operations with the Arena model described above alone, we are interesting to bring an optimization module to our implantation. More precisely, we want to change periodically the block lengths according to some optimization criteria. To achieve this, we have developed an optimization function that takes as parameters a set of variables describing the state of the network at a given time and that returns the new block lengths, resulting from the resolution of the block length optimization model. This model is solved using Cplex and the function is written in C++. With this C++ optimization function we have created a "dll" library that we call in Arena through VBA. In fact, it is possible in Arena to incorporate VBA code that is called periodically during the execution of the Arena model. With VBA code, we can access variables of the Arena model and change their values. In our situation, we access different variables describing the state of the network. With those variables, we build the block length optimization model that uses Cplex to compute new values for the block length, and then we change the Arena model variables corresponding to the block length before quitting the VBA code.

8. General assumptions regarding the model

(1) When we fill a block to its full capacity with empty cars, we suppose that the empty cars we use have a waybill that has the current terminal as the car origin and the block destination as the car destination. By doing this, the empty car becomes available at its destination and can be used to load containers to be a part of a block, or can be used empty once again to fill a block to its capacity.

(2) There are in fact much more than two car types, but to simplify a little bit the model we classify them according to the maximum weight they can handle with respect to length. A car can handle 20' and 40' containers. The configuration of a car is as follows: on top only may be put a 40' container; on the bottom there may be two 20' containers or one 40' container. 20' containers always go in pairs. On two empties 20' or one 40' empty one can never put a loaded container. But on one empty 20' container and on one loaded 20' container, or on two loaded 20' containers or on one 40' loaded container one can put a loaded 40' container. Of course, the configuration of the car must respect the weight limitation of the car.

(3) According to the CN, an empty 20' container weights about 5000 pounds, so we suppose that a loaded 40' container weights (8000 to) 9000 pounds.

(4) While making the blocks, we need to respect a block length. We suppose that this length is the length of all the cars making the block. We suppose the length of a car is 68'.

(5) We classify the containers into eight types. The rationale behind this classification is that while it is difficult (sometimes impossible) to combine loaded containers with respect to the weight limitation of the cars, it is easier to combine heavy loaded containers with "not so heavy" other loaded containers. For example, we cannot put a very heavy forty feet container with two very heavy twenty feet containers, but this configuration is possible on a high capacity car if we choose a heavy loaded forty feet container (60000 to 65000 pounds) with two "light" loaded twenty feet containers (0 to 35000 pounds); this combination gives us a maximum weight of 161000 pounds.

(6) According to CN, only one waybill is possible between any pair of origin/destination in the network, a waybill being defined as the succession of blocks that carry the demand from its origin to its destination. We have generated all the possible waybills from the blocks. Sometimes, several waybills could carry a container from its origin and its destination. In this case, only the shortest one is considered. In a future scenario, the actual waybills will be included.